***ALMA MATER STUDIORUM – UNIVERSITÀ DI BOLOGNA***

*DOTTORATO DI RICERCA*

# in BIODIVERSITA' ED EVOLUZIONE

**Ciclo XXI°**

**Settore/i scientifico disciplinari di afferenza: BIO/05**

**TITOLO TESI**

# "Mitochondrial Genomics: structure, inheritance and phylogenetic utility of the Mitochondrial genome in Bivalvia and Insecta"

Presentata da: Dr. Andrea Ricci

**Coordinatore Dottorato** — **Prof. Giovanni Cristofolini**

**Relatore** — **Dr. Marco Passamonti**

Esame finale anno 2009

1. FOREWORD ..... 2
2. INTRODUCTION ..... 2
2.1 THE MITOCHONDRIAL DNA ..... 2
2.2 ANIMAL MITOCHONDRIAL DNA: DISTINCTIVE FEATURES ..... 4
2.2.1 gene content ..... 6
2.2.2 genome architecture ..... 8
2.2.3 The asymmetric distribution of genes ..... 9
2.3 THE MITOCHONDRIAL GENOME IN BIVALVE MOLLUSKS ..... 11
2.4 BIVALVIA AND DOUBLY UNIPARENTAL INHERITANCE (DUI) ..... 13
2.4.1 DUI and sex-ratio ..... 15
2.4.2 DUI, selection and molecular evolution of gender-mitochondrial genomes ..... 17
2.4.3 “Masculinization” or “role reversal-event” and its implications for Mitochondrial dna recombination ..... 18
2.4.4 DUI distribution and phylogeny of paternal and maternal genomes. ..... 21
2.4.5 The mytilid *Musculista senhousia*: an unusual case of gender-associated mitochondrial dna heteroplasmy. ..... 23
2.5 THE MITOCHONDRIAL GENOME IN INSECTA ..... 25
2.5.1 The paleozoic diversification of the insect orders and the “ORTHOPTEROIDS PROBLEM” ..... 29
3 RESULTS AND DISCUSSION ..... 30
3.1 MOLLUSCA BIVALVIA ..... 30
3.1.1 Overall gene order ..... 30
3.1.2 Nucleotide composition ..... 31
3.1.3 Ribosomal and transfer RNA genes ..... 32
3.1.4 Gene junctions ..... 32
3.1.5 Gene content ..... 33
3.1.6 Comparative sequence analysis ..... 35
3.1.7 Codon usage ..... 36
3.1.8 Large Unassigned region (LUR) and sex-linked mt-DNA transmission ..... 38
3.2 PHASMATODEA ..... 40
3.2.1 Results ..... 40
3.2.2 Phylogenetic analyses and effect of data manipulation ..... 41
3.2.3 Discussion ..... 43
3.2.3.1 Methodological effects of various approaches to phylogenetic reconstruction ..... 45
4. Materials and methods ..... 46
4.1 SECTION-1: MOLLUSCA BIVALVIA ..... 46
4.1.1 Sample collection ..... 46
4.1.2 PCR amplifications, sequencing and annotation ..... 46
4.1.3 Annotation, alignment and phylogenetic analyses ..... 47
4.2 SECTION-2: PHASMATODEA ..... 48
4.2.1 Sample collection ..... 48
4.2.2 PCR amplifications and sequencing ..... 48
4.2.3 Annotation, alignment and phylogenetic analyses ..... 50
4.2.4 Phylogenetic analyses and data partitioning ..... 50
5. REFERENCES ..... 52
6. TABLES ..... 78
7. FIGURES ..... 90
8. APPENDIX ..... 100

## 1. FOREWORD

This PhD Thesis is the result of my research activity in the last three years. My main research interest was centered on the evolution of mitochondrial genome (mtDNA), and on its usefulness as a phylogeographic and phylogenetic marker at different taxonomic levels in different taxa of Metazoa. From a methodological standpoint, my main effort was dedicated to the sequencing of complete mitochondrial genomes, and the approach to whole-genome sequencing was based on the application of Long-PCR and shotgun sequences. Moreover, this research project is a part of a bigger sequencing project of mtDNAs in many different Metazoans' taxa, and I mostly dedicated myself to sequence and analyze mtDNAs in selected taxa of bivalves and hexapods (Insecta).
Sequences of bivalve mtDNAs are particularly limited, and my study contributed to extend the sampling. Moreover, I used the bivalve *Musculista senhousia* as model taxon to investigate the molecular mechanisms and the evolutionary significance of their aberrant mode of mitochondrial inheritance (Doubly Uniparental Inheritance, see below).
In Insects, I focused my attention on the Genus *Bacillus* (Insecta Phasmida). A detailed phylogenetic analysis was performed in order to assess phylogenetic relationships within the genus, and to investigate the placement of Phasmida in the phylogenetic tree of Insecta. The main goal of this part of my study was to add to the taxonomic coverage of sequenced mtDNAs in basal insects, which were only partially analyzed.

## 2. INTRODUCTION

### *2.1 THE MITOCHONDRIAL DNA*

Mitochondria are known for their "cellular power plants" role, that is the production of ATP through oxidative phosphorylation. Such a role is unquestionably of great interest, but many other important features and functions are coming to light as the numerous researches proceed. As a matter of fact, there are strong evidences about an involvement of mitochondria in many biological processes of the eukaryotic cell life. Mitochondria have a central role in the control of cellular cycle (in particular cell growth and death; see McBride et al., 2006 for a review) and are responsible for several aspects in cell signaling, fertilization, development, differentiation, ageing, apoptosis and even sex determination, just to name some (Lenaz, 1998; Werren and Beukeboom, 1998; Wang, 2001; Brookes et al., 2002; Van Blerkom and Davis, 2007). Moreover, mitochondria are implicated

in human diseases such as miopathy, Kearns-Sayre syndrome, MELAS syndrome, Parkinson's disease and other neurodegenerative diseases (see Schapira, 2006 for a review). It is clear that the already strong interest around these organelles is destined to grow.

Metazoan mitochondrial genome (mtDNA) is a good model system for evolutionary genomic studies and the availability of more than 1000 sequences (at present) provides an almost unique opportunity to decode mechanisms of genome evolution over a large phylogenetic range.

The mitochondrial genome (mtDNA) of Metazoa is the marker of choice for phylogenetic reconstruction at several taxonomic levels, from populations to Phyla, and has been widely used to resolve taxonomic controversies. Indeed, the small size of the molecule, its abundance in animal tissues, the strict orthology of encoded genes, the presence of genes/regions evolving at different rates, its uniparental inheritance (with exceptions to the rule) and the absence (or very low level) of recombination (Elson and Lightowlers, 2006) make this molecule a reliable and easy-to-use phylogenetic marker.

Besides its use as a phylogenetic marker, the mtDNA represents a 'full' genome and the availability of complete sequences provides a unique, largely unexplored, opportunity to decode the mechanisms underlying genomic evolution in a phylogenetic framework. Several structural genomic features (for example genome size, gene content, gene order, compositional features, nucleotide substitution rate, repeated sequences, non-coding sequences, secondary structure of the encoded RNA) can be systematically and quite easily investigated in the small mitochondrial genome. These features allow both the description of evolutionary trends in phylogenetically distant organisms and the identification of differences in functional constraints that might account for structural differences. Actually, differences in structural genomic features generally reflect differences in functional and evolutionary constraints. At present, only the rate and pattern of nucleotide substitutions have been widely analysed in mtDNA: in vertebrates, a correlation has been demonstrated between the asymmetrical process of genome replication and mutation pattern, substitution rate and compositional asymmetry, given that all these parameters are proportional to the time spent by the H strand in single-strand status, and vary with respect to genomic position (Bielawski and Gold, 2002; Faith and Pollock, 2003; Broughton and Reneau, 2006). The evolutionary and functional meanings of other mt features still await more detailed and systematic analyses, although some general hypotheses can be put forward. For instance, gene content is certainly affected by the ability of exchange of genetic material between the mitochondrial and nuclear compartments, the permeability and/or the presence of specific carriers on the mitochondrial membranes, gene dispensability and the difference in multimeric structure of the respiratory chain complexes between organisms.

The secondary structure and size of transfer RNAs (tRNAs) and rRNAs are related to peculiarities of the mitochondrial translational apparatus, as demonstrated in nematodes by the concomitant unusual structure of mt tRNA, ribosomal RNAs (rRNA) and elongation factors (Okimoto and Wolstenholme, 1990; Okimoto et al., 1994; Sakurai et al., 2001, 2006; Ohtsuki et al., 2002). The number, size and location of non-coding regions is mostly related to the presence of replication and transcription regulatory signals, although the overlap between regulatory signals and coding sequences cannot be excluded in the compact metazoan mtDNA (Valverde et al., 1994; Peleg et al., 2004). Finally, gene arrangement is affected by the transcription mechanism, for example by the need to co-regulate the expression of some genes or the required stoichiometry of the gene products. Gene order changes can also originate from processes strictly related to replication, such as tandem duplications of genomic segments due to slipped-strand mispairing or imprecise termination of replication (Boore, 2000).

Although mtDNA rearrangements may have profound functional implications on gene expression and genome replication, they have mainly been used to extract phylogenetic information, rather than to extrapolate data on the transcription or replication mode: one of the few exceptions is the prediction of the functionality of the duplicated control region (CR) in snake mtDNAs based on the analysis of the variability of compositional skew along the genome (Jiang et al., 2007).

### *2.2 ANIMAL MITOCHONDRIAL DNA: DISTINCTIVE FEATURES*

In animals, mtDNA is generally a small (15-20 kb) genome containing 37 genes. Although much larger mitochondrial genomes have occasionally been found, these are the products of duplications of portions of the mtDNA rather than variation in gene content (Boyce et al., 1989; Fuller and Zouros, 1993). The typical gene complement encodes 13 protein subunits of the enzymes of oxidative phosphorylation, the two rRNAs of the mitochondrial ribosome, and the 22 tRNAs necessary for the translation of the proteins encoded by mtDNA.

Among multicellular animals this is nearly always a closed-circular molecule; only the cnidarian classes Cubozoa, Scyphozoa and Hydrozoa have been found to have linear mtDNA chromosomes (Bridge et al., 1992).

All of the 37 genes typically found in animal mtDNA have homologs in the mtDNAs of plants, fungi and/or protists (Boore and Brown, 1994; Levings and Vasil, 1995; Gray et al., 1998; Paquin et al., 1997). There is also generally a single large non-coding region which is known to contain controlling elements for replication and transcription. Whether these ‘control regions’ (assumed to be the largest non-coding segment) are homologous between distantly related animals or,

alternatively, have arisen from various non-coding sequences independently in separate evolutionary lineages is often unclear since they share no sequence similarity except among closely related animals.

In some mtDNAs all genes are transcribed from one strand, whereas in others the genes are distributed between the two strands.

In the cases where it has been studied, each strand is transcribed as a single large polycistron which is post-transcriptionally processed into gene-specific messages. Transfer RNA genes, whose secondary structures are thought to signal cleavage (Ojala et al., 1981), punctuate the polycistron of some mtDNAs, but others have tRNA genes clustered so as to preclude such a mechanism. Although potential secondary structures have been proposed as substitutes at some of these gene junctions, no obvious secondary structures can be identified for most cases. It is possible that the 'polycistron model' may not be appropriate for many of the animals whose mitochondrial gene expression remains unstudied.

Concerning the distribution of the completely sequenced mtDNAs among the major metazoan groups, the most represented one is the subphylum Vertebrata followed by the phylum Arthropoda (70% and 13% of the total data set, respectively), and the most abundant species are neopterygian fishes and eutherian mammals (31% and 16% of the total data set, respectively). A total of 14 phyla are represented by only 1 or 2 sequences, and 5 of them include only partial mtDNAs (Myzostomida, Phoronida, Nemertea, Pogonophora and Sipuncula), thus indicating that many gaps in the taxonomic sample of metazoan mtDNAs need to be filled.

At present the size of the completely sequenced mtDNAs ranges from 32.115 bp of *Placopecten magellanicus* (Mollusca Bivalvia) to 11.423 bp of *Paraspadella gotoi* (Chaetognatha), but one of two circular mt molecules of the rotifer *Brachionus plicatilis* is slightly smaller (11.153 bp; Suga et al., 2008).

Moreover the size of the mt genome varies remarkably among the major metazoan groups, and nonetheless it can be highly variable even within the same group. Chordata, Echinodermata, Arthropoda and Platyhelminthes are characterized by a more stable mtDNA size among the most sampled metazoans. On the contrary, Mollusca, enoplean Nematoda and Porifera exhibit a strong heterogeneity in genome size. Variation in the length of the CR accounts for much mtDNA size variability, because of repeated sequences are often present. Anyway, the largest differences in genome size are due to segmental duplication followed by several rearrangements, with consequent presence of duplicated genes or pseudogenes and increased length of non-coding sequences.

### *2.2.1 GENE CONTENT*

Although the most frequent mt gene content consists of 37 genes, nonetheless the gene number ranges from 14 genes in two Chaetognatha species to 53 in *Metaseiulus occidentalis* (Arthropoda, Chelicerata; Helfenbein et al., 2004; Papillon et al., 2004; Faure and Casanova, 2006; Jeyaprakash and Hoy, 2007). This variability in gene content is mostly due to tRNA genes, and rarely concerns protein-coding and rRNA genes. Although the current view of metazoan phylogeny is partially resolved and the mtDNA sequences are extremely limited for many taxa, it is clear that the overall mtDNA plasticity in Metazoa is higher than previously thought. Mainly, this plasticity is dued to rRNA genes because of their number is highly variable both among and within taxa.

As a consequence, that metazoan mitochondria would appear as a quite flexible system with respect to both their ability to import tRNAs from the cytoplasm and the tolerance of the mtDNA towards the acquisition or loss of functional tRNA genes. Tranfer-RNAs are a gene category highly 'dispensable' in the mitochondrial genome. In fact, their number depends not only on the mt genetic code but also on the evolutionary history of the taxa. Moreover unknown factors might be related to the mtDNA's ability to acquire/lose genetic material and to the capacity of the organelle to import and use nuclear-encoded tRNA molecules.

The case of tRNA-Met rapresents an interesting example of loss/acquisition of tRNA genes. In fact, two distinct genes for an initiator and an elongator tRNA-Met, both with CAU anticodon, have been identified only in the phylum Placozoa, the basal metazoan lineage (Dellaporta et al., 2006; Signorovitch et al., 2007) and in Porifera Demospongiae. Both tunicates and some bivalves rapresent an exception, since an unusual tRNA-Met(UAU) has been recognized as an elongator tRNA-Met additional to the common tRNA-Met(CAU) sharing with the others metazoan lineages (Hoffmann et al., 1992; Beagley et al., 1999; Yokobori et al., 1999, 2003, 2005; Gissi et al., 2004).

This peculiar phylogenetic distribution suggests that the elongator and initiator tRNAs-Met have been lost early in metazoan diversification, and have been re-acquired independently in the two distant lineages of mollusc bivalves and tunicates.

Additional protein-coding genes are present only in the mtDNA of non-bilaterians animals, such as Cnidaria and Porifera. Although some of these genes are open reading frames (ORFs) with no similarity to known proteins, it has been observed that the additional proteins of Cnidaria are mostly involved in DNA interaction, whereas those of Porifera have metabolic roles (Flot and Tillier, 2007) as they are also encoded by the closest relatives' mtDNA, *i.e* the atp9 and *tatC* genes identified in Porifera are also encoded by some protists' mtDNA (Yen et al., 2002; Burger et al., 2003; Dellaporta et al., 2006), indicating that this is one of the several primitive features retained by

Porifera (Haen et al., 2007).

The sporadic presence/absence of additional genes in non-bilaterian species indicates a tendency of their mtDNAs to acquire or loose genetic material, sometimes through the activity of mobile elements. Indeed, Hexacorallia possess one or two introns of group I, probably still active as shown by the tendency of the nad5 intron to gain genes from the rest of the genome (Medina et al., 2006; Brugler and France, 2007; Sinniger et al., 2007), whereas in the sponge *Amphimedon queenslandica* the *atp9* gene has been transferred to the nuclear genome, probably through a transposon-mediated translocation, as the functional nuclear *atp9* gene is flanked by inverted repeats (Erpenbeck et al., 2007).

Finally, introns of group I have been identified only in Cnidaria and Porifera, whereas surprisingly a group II intron has been found in the annelid *Nephtys sp*., probably as result of a recent horizontal gene transfer from a bacterial or viral vector (Valles et al., 2008).

With respect to the other mt-encoded genes, the two rRNA genes are always mitochondrially encoded, and their duplication is very rare, having been observed in only four species (*rrnS* is duplicated in the bivalve *Crassostrea gigas* and in distinct haplotypes of the pillbug nematode *Thaumamermis cosgrovei* (Milbury and Gaffney, 2005; Tang and Hyman, 2007); *rrnL* is duplicated in the chigger mite *Leptotrombidium pallidum* (Shao et al., 2005b); both *rrnS* and *rrnL* are duplicated in the large mt haplotype of the nematode *Strelkovimermis spiculatus*.

Loss or acquisition of the protein-coding genes are also quite rare. Nonetheless, the actual gene loss remains sometimes ambiguous because of either uncertainty in annotation or incomplete mtDNA sequences (i.e. the absence of *nad3* and *nad6* in the mite *Metaseiulus occidentalis*, Jeyaprakash and Hoy, 2007, and the absence of *atp8* and *nad6* in two Hexactinellida sponges, Haen et al., 2007, respectively).

Only a few cases of sporadic gene loss have been reported in vertebrates. Actually, the loss of *nad6* has been described in the neopterygian *Choinodraco rastrospinosus*, and reported as a common feature of all Antarctic Nothothenioidei, therefore linked to the actual loss of protein function (Papetti et al., 2007).

An interesting observation is that genes encoding for subunits of the adenosine-triphosphate synthase complex (ATPase) showed a significant tendency to be lost or acquired by animal mtDNA. Actually, most Porifera mtDNAs have the *atp9* gene, while the *atp6* gene is absent in all Chaetognatha.

Interestingly, in several phylogenetically distant groups, that is in Chaetognatha, Rotifera, most Mollusca Bivalvia (13 of the 19 sampled mtDNAs), Nematoda and Platyhelminthes, the *atp8* gene is absent or highly modified. Indeed, putative *atp8* genes have been found in one rotifer (Steinauer

et al., 2005) and in the nematode *Trichinella spiralis* (Lavrov and Brown, 2001a), but none of these proteins contains the MPQL amino acid signature conserved at the N-terminal of other metazoan ATP8 (Gissi et al., 2004). It should be stressed that some features have hampered the annotation of *atp8* in several cases; the ATP8 protein is actually characterized by a short and quite variable length, and by a higher conservation of its secondary structure, coupled with a high variability of the primary sequence (Papakonstantinou et al., 1996a, b; Gray, 1999). Indeed, among bivalves, *atp8* gene has been annotated in *Hiatella* (Heterochonchia; Dreyer and Steiner, 2006) and *Lampsilis* (Palaeoheterodonta; Serb and Lydeard, 2003) only, but Gissi et al., (2008) have identified this gene also in the M and F types of both *Venerupis* and *Inversidens* species. Similarly, *atp8* was not annotated in the first published tunicate mtDNA of *Halocynthia roretzi* (Yokobori et al., 1999), although a very short form of ATP8 is actually mitochondrially encoded in all tunicates (Gissi and Pesole, 2003; Yokobori et al., 2005; Iannelli et al., 2007a). Despite these annotation difficulties and, because of *atp8* is absent in several distant metazoan lineages, it has been hypothesized that *atp8* is a 'dispensable' gene in the mtDNA, and that its loss (occurred several times) usually is coupled with an increase of the mt substitution rate. Nonetheless, further investigations will be necessary to understand whether this loss is related to the transfer of the gene to the nuclear genome, or to differences in the ATP synthase structure between taxa.

### *2.2.2 GENOME ARCHITECTURE*

Gissi et al. (2008) defined the mtDNA "AR" index as the order of the entire set of functional mt-encoded genes, included duplicated and unusual mt genes. This index is strictly related to both gene content and gene order, and a difference in "AR" means a difference in gene content and/or gene order.

They estimated the variability in mtDNA "AR" for each major metazoan group by the 'AR rate', given by

$$(NAR - 1)/(N_{mtDNA} - 1)$$

where NAR and $N_{mtDNA}$ are the number of different ARs and the number of completely sequenced mtDNAs of that taxa, respectively. In this regard, the AR rate ranges between 0 (no AR variability) and 100 (all mtDNAs have a different AR).

Actually, they observed AR values higher than 70 in nine phyla (Hemichordata, Annellida, Brachiopoda, Chaetognatha, Bryozoa, Entoprocta, Rotifera, Mollusca and Porifera), as well as in subgroups belonging to phyla with good conservation of mtDNA AR: Tunicata within Chordata,

Myriapoda within Arthropoda, and Enoplea within Nematoda.

Vertebrates are the sub-phylum with the lowest variability in mtDNA AR, when compared to other groups of similar taxonomic rank. This would suggest that the almost frozen mtDNA gene order and content, ascribed to Vertebrates, is an exception, rather than a rule.

By excluding tRNAs, which are often described as highly 'mobile', the number of ARs decreases only slightly in taxa showing the highest AR variability (Tunicata, Hemichordata, Brachiopoda, Chaetognatha, Bryozoa, Rotifera, Mollusca, enoplean Nematoda and Porifera), with the exception of the phylum Annelida, where the extensive rearrangements appear exclusively associated to the translocation of tRNAs.

By contrast, in Vertebrata, Arthropoda and Platyhelminthes (all taxa with moderate variability of genome AR), the exclusion of tRNA genes leads to a large decrease in AR rate. This situation would seem to suggest that different rearrangement mechanisms, characterized by a differential involvement of the tRNA genes, could occur in taxa with a variable versus stable mtDNA AR, or that a 'saturation effect' in the gene order changes has obscured the involvement of tRNA in the highly rearranged mtDNAs.

In general, the observation of a high mtDNA AR variability in several phylogenetically distant groups, indicates that the acceleration rate of genomic rearrangements has occurred independently several times. This fact suggests that the causes of this increased rate could be investigated through the analysis of physiological and metabolic features common to distant taxa sharing the same evolutionary pattern.

### *2.2.3 THE ASYMMETRIC DISTRIBUTION OF GENES*

Animal mtDNA is characterized by an asymmetric distribution of the genes (GSA) between the two strands, allowing the identification of a major and a minor strand depending on the number of encoded genes. The GSA has been quantified by Gissi et al. (2008) as the absolute value of the difference in gene number between the two strands, divided by the total number of genes. The resulting GSA value ranges from 0 to 1, with values close to zero indicating an almost equal number of genes encoded by the two strands, and values higher than 0.5 corresponding to the presence of at least 75% of the total genes on the major strand.

The asymmetric gene distribution is a common feature of almost all metazoans. Actually, quantitative GSA data point out to a large variability of it, ranging from almost 0 to 1, in Arthropoda, Mollusca, enoplean Nematoda and Porifera (Table 1), thus indicating that frequent exchanges of genes between the two mt strands have occurred during the evolution of those

genomes. On the contrary, the GSA is almost invariant and very close to 0.5 in all vertebrates and cephalochordates, thus evidencing rare gene inversions in those groups.

Mitochondrial genomes showing a symmetric gene distribution (GSA less of 0.1, thus 50–55% of the total genes on each strand) are very rare, and this feature has been observed only in 17 phylogenetically distant species belonging to four phyla (a few echinoderms, cephalopods, scaphopods, crustaceans, hexapods, nematodes and sponges).

On the contrary, a wide taxonomic range, 10% of all the complete mtDNA analysed by Gissi et al. (2008), resulted to be characterized by an extreme GSA value, since all the genes are encoded on the same strand and therefore transcribed in the same direction.

This peculiar feature is shared by all, or almost all, mtDNAs of three phyla (Annelida, Brachiopoda and Platelminta), three sub-phyla/classes (Tunicata, Bivalvia and Nematoda Chromadorea) and the non-bilaterians Porifera and Cnidaria. The latter mtDNAs are characterized by a single cluster of co-oriented genes. The few exceptions to 'one-cluster structure' have been found in Porifera, Cnidaria and Bivalvia, consisting in two clusters of co-oriented genes (one for each strand). Similarly, the mtDNAs of non-bivalve molluscs are characterized by a structure with one-, two- or four-exact or defective clusters of co-oriented genes (defective clusters are defined as co-oriented gene clusters interrupted by the presence of one or two adjacent genes located on the opposite strand).

Thus, these features (*i.e.* the most extreme GSA value) have led to the hypothesis that this gene organization was likely ancestral. Moreover the exceptions to the rule of 'one coding-strand' found in non-bilaterians and lophotrochozoans (especially molluscs) mostly consist of mtDNAs with a few perfect or defective clusters of co-oriented genes, which would suggest that the ancestral gene organization was then modifed by the inversion of large mtDNA segments, rather than by strand translocation of single genes.

The maximum variability in GSA (ranging from 0 to about 1) in two large ecdysozoan phyla (Nematoda and Arthropoda) is also in accordance with the theory of an ancestral 'one coding-strand' mtDNA structure that was slowly modified during metazoan diversification.

Finally, it is intriguing that five of the eight above reported taxa with only one coding strand are also characterized by a strong variability of genome AR (Tunicata, Annelida, Brachiopoda, Bivalvia and Porifera), suggesting a possible correlation between these two features and that the same mechanism could be responsible for both the presence of a single cluster of co-oriented genes and a strong variability in gene number/order.

## *2.3 THE MITOCHONDRIAL GENOME IN BIVALVE MOLLUSKS*

Mollusca are the second largest animal phylum. Currently, out of 42 complete mollusk mtDNAs available in Genebank, 13 are from bivalves, including *Mytilus edulis*, *M. trossulus*, *M. galloprovincialis*, *Venerupis philippinarum*, *Inversidens japanensis* (all male and female mitotypes), *Lampsilis ornata*, *Placopecten magellanicus*, as well as three from oyster species: *Crassostrea gigas*, *C. hongkonghensis* and *C. virginica.*

Among sequenced mtDNAs, there are very few shared gene boundaries, and gene translocations appear across all gene classes (protein-coding, *tRNA* and *rRNA*). Because of their high variability, Bivalvia may provide informations that can be used to develop models for the mechanism involved in gene rearrengement, replication and regulation; in fact, gene duplication and/or loss is present in almost every taxon in which a complete mt genome is available. For example, *rrnS* is duplicated in both *Crassostrea gigas* and *C. virginica* while in *C. hongkongensis* the duplication is not present; *tRNA* genes have been duplicated in *Mytilus* (*trnM*), *Venerupis* (*trnM* and *trnV*), as well as in the three *Crassostrea* taxa (*trnM*, *trnL* and *trnS*) while *trnQ* is only duplicated in *Crassostrea hongkongensis*. In this latter, *trnC* and *trnN* genes are absent, a phenomenon not reported before for any other mollusk species. Moreover, the female mitotype of *Venerupis philippinarum* has a second tandem copy of *cox2* with different length.

Hoffman, Boore, and Brown (1992) were the first to report that the protein-coding gene *atp8* is absent in *Mytilus*, a condition that was only previously known to occur in nematodes (Okimoto et al.1991). Subsequent studies have shown that other bivalves also lack this gene, including *Crassostrea* and *Venerupis*. Although *atp8* is present in the *Inversidens* F-mitotype, it appears to be nonfunctional due to the presence of several stop codons after the first 31 amino acids, while, quite interestingly, it does not appear to be an identifiable remnant of this gene in the male mitotype. In contrast, the unionid, *Lampsilis* possesses a complete copy of *atp8*. However, this copy encodes a protein that is longer than the ‘‘typical’’ ATP8 by 13 extra amino acids on the C-terminus and it has not been determined whether this gene is functional in *Lampsilis*.

In this regard, not much is known about the mechanisms driving new mitochondrial gene rearrangements. The duplication-random loss model (Moritz et al., 1987) and the duplication-non-random loss model (Lavrov et al., 2002) may account for larger rearrangements. According to the duplication-random loss model, partial duplication of mtDNA caused by errors in replication, such as erroneous initiation or termination, or strand slippage and mispairing, followed by the loss of one copy of each duplicated gene, are taken into account to explain rearrangements (Moritz et al., 1987; Madsen et al., 1993; Moritz et al., 1993; Macey et al., 1997; Boore, 2000).

After duplication, it is commonly assumed that the loss of one of the two copies of each duplicated

gene happens at random. However, according to the duplication-non-random loss model, the destiny of each gene copy in the duplicated region is predetermined by its transcriptional polarity and location in the genome (Lavrov et al., 2002). All genes having one polarity would be lost from one genome copy, and all genes having the opposite polarity would be lost from the other. Two other less discussed models are the dimerization of two mt genomes and 'illegitimate' recombination between mt genomes (Boore 2000). The general lack of knowledge for the mechanisms involved in gene translocation and genome rearrangement affects the interpretation of gene order data. Bivalves, with their great diversity, ancient lineages that span the Paleozoic (Skelton and Benton 1993), and the observed variation in genome content should eventually provide a good system for modeling the mechanisms of gene movement.

With regard to this, the available mt gene order data do not support a particular model of gene rearrangement for bivalves, but it does provide tantalizing clues. Multiple gene copies exist in extant taxa, but the majority of gene duplications are not in tandem. This may be because either the duplication did not form in tandem or the gene translocation occurred subsequently to the duplication event. Sampling mitochondrial genome sequences of more closely related bivalves may provide insight to the mechanism of gene movement by limiting the number of gene translocations. Bivalves provide a unique opportunity to formulate models by providing historical ''snapshots'' of gene movement, as duplications are predicted to be intermediate steps in mt genome rearrangements. Also, because of their unique mode of mtDNA inheritance (DUI: doubly uniparental inheritance, see below), bivalves may have mechanisms for gene translocation not present in other mollusks. In fact, the occurrence of doubly uniparental inheritance found in some bivalve lineages may also be involved in gene translocation, gene duplication/loss events, and recombination.

Bivalves possess two distinct mtDNA types, that is a female (F) and a male (M) type differentially transmitted to the progeny through DUI (Zouros et al., 1994). In DUI species, the extent of genome rearrangement between the two gender-specific mtDNAs is quite variable, and it ranges from the translocation of several genes in *Inversidens japanensis* (no gene inversion), to differences in gene content in *V. philippinarum* and in some isoforms of the *Mytilus* genus (Mizi et al., 2005; Breton et al., 2006; Zbawicka et al., 2007). The gene order is identical in the three analyzed species of *Mytilus*, also described as members of the same species complex (Gossling, 1992), and duplication of the CR and some adjacent genes has been sporadically found only in some M and 'recently masculinized' mtDNA types of *Mytilus* (Mizi et al., 2005; Breton et al., 2006). In addition to these features, species with DUI show an accelerated nucleotide substitution rate compared to other animals.

### *2.4 BIVALVIA AND DOUBLY UNIPARENTAL INHERITANCE (DUI)*

Given the universality of uniparental transmission of organelle DNA, the finding that in the mussel *Mytilus edulis* part of the progeny carried mtDNA of both parents was a big surprise (Zouros et al., 1992). In a previous study, Fisher and Skibinski (1990) observed that heteroplasmy was common in *Mytilus edulis*, and that it was higher among males than females. As above mentioned, heteroplasmy was already known in many animals, but *Mytilus* one was linked to sex.

In another study, focused on the amount of mtDNA divergence, an individual mussel resulted to be heteroplasmic, but no link with its sex was examined (Hoeh et al., 1991). Overall, they found that the value of mtDNA divergence was unusually large (more than 20%), and because this amount could not be explained by gradual accumulation of mutations in the two mt genomes in heteroplasmy, "Hoeh et al" provided a reasonable explanation assuming that the two molecules came in the same individual from two preexisting and differentiating (for a long time) lineages.

All these observations suggested that the only obvious way two differentiated mtDNAs could come in the same individual were through egg and sperm, *i.e.* through a biparental inheritance.

Skibinski et al. (1994a, 1994b) and Zouros et al. (1994a, 1994b) were the first to resolve the above mentioned puzzle. These authors showed that heteroplasmy in mussels was associated to a gender-associated mtDNA transmission system that required distinct paternal and maternal mitotypes. In more detail, Skibinski *et al.* analyzed somatic tissues and gonads separately in both male and female individuals. In male soma, the observed mtDNA was the females one (referred to for convenience as "F" or F-type). but males also had in their somatic tissues another mtDNA-type that was not normally found in females (the so called "M" or "M-type"). The female gonad carried the F-type molecule, but the testicles carried, very surprisingly, only the M-type molecule.

To explain such pattern of sex-linked heteroplasmy, Zouros *et al*. (1994a, 1994b) analyzed the progeny of *M. edulis* crosses and they observed that all females inherited exclusively the mother's mtDNA, with almost no exception. Conversely, all the males inherited two genomes: the father's and the mother's ones. Moreover the father's maternal genome was not passed to any offspring, male or female.

This anomalous mtDNA transmission system seems at first to violate the rule of uniparental inheritance because males receive mtDNA from both parents. Consequently their mtDNA inheritance is indeed biparental. However, the apparent violation to uniparental inheritance vanish because the two mtDNAs independently segregate in mussels: the F-type is transmitted from mothers to both sexes, but is then transmitted to the next generation only through females, just as in the standard mode of maternal inheritance, while the M-type is transmitted from father to sons only. Thus both genomes are uniparentally transmitted, each through one sex. For each genome the rule

of uniparental transmission does apply, but simply operates through two routes. For this reason this mtDNA inheritance has been called Doubly Uniparental Inheritance (DUI).

DUI challenges the traditional view of the strict maternal inheritance (SMI) of mtDNA. A striking difference between SMI and DUI systems is that the latter allows selection to act directly on the male mitochondrial genome; unlike most animals, male mussels do not represent an evolutionary dead-end for mitochondrial genomes (Breton et al., 2007 and references therein).

Moreover, this system provides a unique opportunity to study mitochondrial behavior from a different and privileged point of view, because a non-traditional model could be very useful in answering many questions that cannot be effectively solved by studying classical maternal mitochondrial models (Passamonti and Ghiselli, 2009).

Since then, DUI was documented in other (but not all) bivalves: figure xx shows the 36 bivalve species that evidenced DUI, according to the phylogeny of Bivalvia by Giribet and Wheeler (2002). Evidence for DUI is obtained from the literature for 31 species (Fig. xx). For other five species, such evidence is based on sequences retrieved from GenBank (sequences deposited by Okazaki, M., Shikatani, M., Nishida, M., and R. Ueshima). These sequences were derived from female and male gonads and, although very similar within genders, they differ between genders by more than 10%, a difference that is not expected from homospecific sequences of species with uniparental inheritance. The 36 bivalve species with DUI belong to seven families and five superfamilies. Given the broad taxonomic distribution of DUI, it was therefore proposed that this system evolved once in an ancestral bivalve lineage and was subsequently lost in some descendants. Unfortunately, there is presently no information about DUI in any species of the basal superfamilies Solemyoidea or Nuculoidea (see below).

The transition from strict maternal inheritance (SMI) to DUI probably involved a modification of the recognition system of sperm mitochondria by eggs, and a specific mechanism ensuring a father-to-son transmission of M mtDNA (Zouros 2000). Empirical evidences for such mechanisms came from studies on *Mytilus* embryos. Typically, the fate of sperm mitochondria depends on whether the embryo is destined to develop into a female or a male (Saavedra et al., 1997; Sutherland et al., 1998; Cao et al., 2004; Cogswell et al., 2006; Venetis et al., 2006). For example, Sutherland et al. (2008) found that during fertilization, all eggs receive sperm mitochondria, which are eliminated or drastically reduced within 24–48 hrs in female embryos. Recent epifluorescence-based observations of embryos destined to become males demonstrated that sperm mitochondria tend to aggregate in a single blastomere that is thought to give rise to the male germ line. By contrast, in embryos destined to become female, sperm mitochondria are randomly dispersed among blastomeres (Cao et al., 2004; Cogswell et al., 2006). This sex-specific difference in the embryonic aggregation of M versus

F genomes, which is one factor responsible for the tissue specific differences in ratios of M and F genomes, appears to be dependent on the action of microtubules (Obata and Komaru, 2005; Yaffe, 1999). In contrast, a different mechanism has been described for *Crassostrea gigas*, which is a strict SMI bivalve species (Obata et al., 2008).

These studies suggest that heteroplasmy is the initial state in the early development of mytilid embryos. In the family Mytilidae, females normally shift from being heteroplasmic zygotes to essentially homoplasmic adults. Mature male mytilids contain different ratios of F and M genomes in their tissues (*i.e.* testes contain predominantly M genome, somatic tissues contain predominantly F genome) (Garrido-Ramos et al., 1998; Dalziel and Stewart, 2002). By contrast, the venerid clam *Venerupis philippinarum* has a strong predominance of M mtDNAs in somatic tissues (Passamonti and Scali, 2001). Taken together, these results indicate that mechanisms making sperm mitochondria and M genome into the male gonad are not always perfect (Zouros 2000). Similarly, mechanisms limiting the segregation of M genomes in developing females are also not perfect. Trace amounts of M genome have been found in tissues of adult females (Garrido-Ramos et al., 1998; Dalziel and Stewart, 2002) and even in the unfertilized eggs of *M. galloprovincialis* (Obata et al., 2006). By contrast, male germ line seems to exclude the F genome and preferentially amplify M genome (Venetis et al., 2006). Indeed, by forcing spawned sperm to swim through a solution of Percoll, and thus minimizing probability of sperm contamination by somatic tissues or cells, Venetis et al. recovered only distinct, paternally transmitted mtDNA genomes in the total DNA extractions from the 'washed' sperm of several *M. galloprovincialis* individuals. This precise male-specific transmission of the M mitochondrial genome is required for the stability of DUI.

### *2.4.1 DUI AND SEX-RATIO*

Another unusual aspect of mussel genetics (and to date this has only been observed in *Mytilus*) is that some females produce female-biased offspring, whereas other females produce male-biased offspring, regardless of which male they mate with.

Within a set of 24 factorial crosses involving 5 females and 5 males, Saavedra et al. (1997) found pogenies that consisted almost entirely of daughters, and progenies in which sons were more than the 85% of progeny. Other crosses had intermediate sex-ratios. The same female parent produced the same sex-ratio in the offspring, regardless the male was mated to, thus the sex-ratio appears to be determined by the female.

Therefore, *Mytilus* females fall into 3 distinct classes: those that predominantly produce daughters, those that predominantly produce sons, and those that produce both sexes in intermediate numbers.

Cao et al. (2004) took advantage of the strong sex-ratio bias that is well established in the marine mussel species *Mytilus edulis – Mytilus galloprovincialis*: using the florescent dye MitoTracker Green F, they stained sperm mitochondria, and such sperm was used to fertilize eggs from females falling into the 3 above mentioned classes. The stained mitochondria were subsequently observed in the cleaving egg. They demonstrated that in embryos destined to become female, sperm mitochondria are randomly dispersed among blastomeres. On the contrary, in embryos destined to become males, sperm mitochondria tend to aggregate in the lager cell resulting from the first or second egg division. This cell, designated "CD" in the 2-cell stage and "D" in the 4-cell stage (Conklin 1897) is the one that is thought to origin male germ line.

In a further study by Cogswell et al. (2006), the authors followed the same experimental approach, but they concentrated their attention on the position of sperm mitochondria in relation to the cleavage furrow (CF) in 2- and 4-cell embryos. Their observations suggested that, after fertilization, sperm mitochondria are released in the area of the egg where division will occur. In embryos from female-biased mothers the sperm were dispersed among the 4 cells moving themselves passively and independently from one another in a apparently random fashion. By contrast, as also observed by Cao et al. (2004), in embryos from male-biased mothers, they observed a preference of mitochondria to migrate to the CD cell or the D cell at the 2- or 4-cell stage, respectively. They also observed that a substantial percentage of these sperm mitochondria are located near to the newly formed membrane of the CD cell. After the second division most of these mitochondria are found in the CF of the cell D (which is a descendant of the CD cell).

Although this provided evidence for the existence of a mechanism that directs sperm mitochondria into the cells that are destined to produce the germ line in male embryos, these observations leave open the question of what causes the different behavior of sperm mitochondria between embryos from female- and male-biased mothers.

Even though there is no direct evidence that the female-driven sex ratio of *Mytilus* is linked to DUI, the fact that both these phenomena, although exceptional, occur in the same organism makes the connection most likely.

Several authors (Zouros, 2000; Saavedra et al., 1997; Cogswell et al., 2006; Kenchington et al., 2002) attempted to develop a reasonable model to explain sex-ratio bias and the coupling of mtDNA inheritance and sex determination in mussels. According to this model, mtDNA inheritance in bivalves with DUI is controlled by a maternally-encoded sex-determining gene, or a gene linked to a sex-determining factor, as has been observed in the basidiomycete fungus *Cryptococcus neoformans* (Xu, 2005).

In more detail, the model proposes that the bias is under the control of the female parent and

suggests that this control is under her nuclear genotype rather than her mitochondrial genotype (Kenchington et al., 2002).

The model involves 3 factors (or 3 sets of factors): W, X and Z: W is located in the outer surface of the sperm mitochondrion and is recognized by X, an egg nuclear-encoded cytoplasmic factor. The model also implies a third DUI-specific factor, Z, which also occurs in the egg cytoplasm.

The X/W complex has been suggested to be part of the 'sperm mitochondria elimination system'. that leads to maternal mtDNA inheritance, maybe a mechanism that could be similar to the ubiquitination system observed in mammals (Kenchington et al., 2002). The Z factor is a one-locus two-allele factor (i.e. the active Z and the inactive z) and its role is to suppress X, thus preventing elimination of sperm mitochondria. Because of this system, females with ZZ genotype will produce all-male offspring, females with Zz will produce intermediate sex-ratio, and females with zz genotype will produce all-female offspring. Therefore, the Z factor, paternal mitochondria and the M genome herein are virtually linked to sex determination.

### *2.4.2 DUI, SELECTION AND MOLECULAR EVOLUTION OF GENDER-MITOCHONDRIAL GENOMES*

An interesting field of analysis concerns the evolutionary forces shaping the mitochondrial genome in DUI species. Actually, mussel mtDNA (both M and F lineages) experienced higher rate of sequence divergence, when compared to other animal taxa; moreover, the M genome evolves more quickly than the F genome.

Actually, the 12 complete or nearly complete F and M mtDNA genomes of DUI species has shown how such species vary from the typical pattern of animal mtDNA gene content: most species [except *Lampsilis* (a unionoidean bivalve)] lost the ATPase subunit 8 (*atp8*) gene, and some have a second *tRNA* gene for methionine (trnM) (e.g. *Mytilus spp*. and female *Venerupis*).

The freshwater mussel *Inversidens japanensis* M and F mitotypes have two gene-order inversions (in both the light and the heavy strands) and two tRNAs (*trnD* and *trnV*) encoded by opposite strands (M. Okazaki and R. Ueshima, unpublished).

The F genome of *Venerupis philippinarum* contains a duplication of the gene for cytochrome c oxidase subunit II (*cox2*) and the M genome contains an extra *trnM* gene, distinguishing the two mtDNAs (M. Okazaki and R. Ueshima, unpublished).

A leading interpretation explaining why M evolves faster than F, and why both of them evolve faster than typical mtDNA is that M and F mtDNAs experience different selective pressures, because of their different "selective arenas" (as proposed by "Stewart et al." 1996). The "Stewart et

al." hypothesis differentiates three selective arenas, namely the somatic cell line, the female germline and the male germline, and it assumes that there might be some tradeoffs in terms of optimal functioning. According to this hypothesis, in DUI species, the M genome has to function in the male germline only, while F genome has to function in soma (of both sexes) as well as in the female germline. The outcome is that the F genome is more constrained and its nucleotide variability should be lower than that of the M genome. This peculiar feature would permit to uncouple the effects of selection for sperm function (acting on the M genome) to the one for somatic metabolic requests (mostly on F).

Studies on *Mytilus*, *Pyganodon* and *Tapes* showed that the M accumulates mutations faster than F one (Liu et al., 1996; Quesada et al., 1998; Passamonti et al., 2003). Higher rate of M mtDNA duplication during spermatogenesis and early male embryo development, free-radical damage to sperm, positive selection, or effects of the smaller population size of the M genome have been proposed to account for this (see Passamonti et al., 2003 for details). Analyses were also performed to evaluate the best molecular evolution model to be applied to *Mytilus* DUI. The neutral model has been rejected (Passamonti et al., 2003), but the nearly neutral theory (Ohta, 1992) has been considered to best fit the observed data (Quesada et al., 1998; Skibinski et al., 1999). However, Skibinski et al. (1999) stated that more information on effective population size is very important to discriminate a nearly neutral situation (Ohta, 1992) from a model of positive selection.

Passamonti et al. (2003), studying 7 protein-coding genes of *Venerupis philippinarum* mitochondrial DNA, evidenced that gene-specific selection has produced two stable and divergent mitochondrial genomes, but they cannot again totally exclude a nearly neutral pattern of evolution.

Another aspect that might deserve mention is that, comparing the same protein-coding genes of *Mytilus* F and M mtDNAs, Breton et al. (2006) found that regions with high or low aminoacid variability are basically the same in both genomes. This was taken as an evidence that "the necessity of evolving in the same nuclear background has apparently forced F and M mtDNA genomes to experience similar selective pressures" (Breton et al., 2006, 2007).

### *2.4.3 "MASCULINIZATION" OR "ROLE REVERSAL-EVENT" AND ITS IMPLICATIONS FOR MITOCHONDRIAL DNA RECOMBINATION*

As above reported, a consequence of DUI is that males are heteroplasmic for the maternal and paternal genome. In spite of this, in marine (but not freshwater) mussels, the fidelity of DUI is sometimes compromised.

Several studies in species of genus *Mytilus* reported males that lacked the M-type genome (Zouros

et al., 1994b; Skibinski et al., 1994b; Hoeh et al., 1997; Saavedra et al., 1997; Rawson and Hilbish, 1995; Quesada et al., 1999; Ladoukakis et al., 2002). Some of these males contained two F-type genomes of which one was predominant in the gonad, an observation which suggested that their paternal genome was F-type. Some other males appeared to contain only one F genome both in their somatic tissues and the gonad, which suggested that either these males failed to inherit a genome from their male parent or that their paternal genome could not be distinguished from the maternal by the assay used.

Those observations provided the basis for the "masculinization" (Hoeh et al., 1997) or "role-reversal" hypothesis (Quesada et al., 1999), according to which an F-type genome may enter the male germ line and become paternally inherited thereafter. Such F genomes are referred to as 'recently-masculinized' M- types ($M^f$) (Zouros, 2000; Hoeh et al., 1997; Quesada et al., 1999; Quesada et al., 2003). Several populations of *Mytilus* are polymorphic for two classes of M mitochondrial genomes: an older, 'standard' M type and a 'recently-masculinized' M type. The genomes of the latter, particularly their protein coding regions, are highly similar to F genomes but they are transmitted as M genomes through sperm.

There are at least two potential explanations for the breakdown of DUI in *Mytilus* spp. According to one view, failure of DUI is associated to hybridization events between *Mytilus* species (coupled with a 4:1 sex ratio bias to females) (Rawson et al., 1996; Saavedra et al., 1996), leading to the disruption of nuclear–cytoplasmic interactions and DUI instability (Curole and Kocher, 2005).

A second explanation came with the discovery of a case of recombination in mussel mtDNA. Actually, Ladoukakis and Zouros (2001) evidenced that intermolecular mitochondrial recombination occurred within the *cox3* gene of male gonad tissues from *Mytilus galloprovincialis* of the Black Sea. They examined heteroplasmic males possessing a typical F genome and $M^f$-masculinized genome; these molecules differed by about 4%. Several different *cox3* gene sequences have been detected in a single individual. No such recombinant haplotypes have been found in other individuals, and it has been speculated that recombination occurred just in the individual tested. It should be noted that both hypotheses are not mutually exclusive; recombination might occur more frequently (or perhaps is more easily detected) in hybrid situations. These findings are highly significant because they provided direct evidence of recombination in animal mtDNA, a much debated subject (reviewed in Rokas et al., 2003). Mitochondrial recombination was later confirmed in other mytilid and venerid species (Passamonti and Scali, 2003; Breton et al., 2006; Rawson 2005; Burzynski et al., 2003; Burzynski et al., 2006).

The fact that occasionally an F genome may assume the role of the M genome and therefore being transmitted trough sperm ("masculinization") seems to point out to the fact that there is no

functional difference between the two genomes (Zouros, 2000). However, studies on M and F genomes evidenced that it might not be the case: the observation that some PCR-amplified main control regions in male gonads of *Mytilus trossulus* were a mosaic of F- and M-like sequences provided a potential link between the homologous recombination of F and M genomes and masculinization (Burzynsky et al., 2003): these recombinant variants (transmitted through sperm like the M genome) showed high coding sequence similarity to the *Mytilus edulis* F genome.

As a consequence, the "Burzynski et al" hypothesis, which based on the fact that occasional invasions of the male transmission route by the F genome were possible through the addition of M control region sequences to the control region of F genomes, may suggest some role of this part on M mtDNA transmission.

The first complete sequence of a recently-masculinized *Mytilus trossulus* mitochondrial genome also indicated that two control regions exist, one male-specific and the other female-specific (Breton et al., 2006). At present, it is not clear whether both potential control regions are functional, whether they act in a tissue-specific manner, or what are their respective roles are in mitotype transmission. Nevertheless, recent evidence of mitochondrial recombination within the control region in male and female *Mytilus trossulus* mussels corroborates the hypothesis that an M-type control region sequence is necessary to confer the paternal role on genomes that are otherwise F-like (Burzynski et al., 2006).

By contrast, in a recent work of Filipowicz et al. (2008) the authors described the rearranged mitochondrial control region in all three European *Mytilus* species. At variance to the earlier above mentioned observations, they suggested that some of the mosaic genomes did not show any gender bias, according to the initial Zouros's supposition (2000).

It is interesting the scenario provided by unionoidean bivalves; the lack of masculinization in freshwater mussels coincides with the presence of a unique M genome-specific 3' extension of the mtDNA-encoded cytochrome c oxidase subunit II gene (M*cox2*) (Curole and Kocher, 2002; Curole and Kocher, 2005; Chakrabarti et al., 2006). This extension is present in all unionoidean M genome *cox2* genes examined to date, including three unionoidean bivalve families (i.e. Hyriidae, Margaritiferidae and Unionidae) (Curole and Kocher, 2002; Curole and Kocher, 2005). Examination of the rates and patterns of substitution suggests that the extension (M*cox2e*) is evolving under relaxed purifying selection relative to the upstream M*cox2* homologous region (M*cox2h*), that is, the region present in both F*cox2* and M*cox2* (Curole and Kocher, 2002; Curole and Kocher, 2005). The M*cox2e* is likely to be the most rapidly evolving mitochondrial domain identified in animals (Curole and Kocher, 2005). Apparently, M*cox2e* is neither present in mytiloid M genomes, nor in other animal mitochondrial genomes (Mizi et al., 2005; Bretone t al., 2006).

A specific function for M*cox2e* has not yet been demonstrated. The *cox2* gene encodes a highly conserved subunit of cytochrome c oxidase, the terminal enzyme of the mitochondrial inner membrane that is responsible for the transfer of electrons from cytochrome c to oxygen (Rawson and Burton, 2006). MCOX2e antibody-based analyses indicated that the extended MCOX2 protein in *Venustaconcha ellipsiformis* (Unionoidea: Unionidae) is predominantly expressed in testes, weakly expressed in other male tissues, and not expressed in female tissues (Chakrabarti et al., 2006). The immunohistochemistry-based localization of MCOX2 to sperm mitochondria combined with the predicted presence of five transmembrane helices in the *V. ellipsiformis* MCOX2e region suggest that the latter is located in the outer and/or inner mitochondrial membrane ((Chakrabarti et al., 2006). These characteristics suggest several potential functions for the MCOX2e; for example, within the sperm mitochondria. Moreover, during spermatogenesis, M*cox2e* expression could be involved in apoptosis, which is an important physiological mechanism that regulates the number of sperm produced (e.g. Chan et al., 2006). One likely developmental function for MCOX2e is that an outer mitochondrial membrane localization could facilitate gender-specific movements of sperm-derived mitochondria within unionoidean embryos in a manner similar to that observed in *Mytilus* (Cao et al., 2004; Cogswell et al., 2006). Although more studies are needed to elucidate the function(s) of the MCOX2e protein, its association with the absence of masculinization in unionoidean bivalves suggests that it has been selected either as a protective mechanism against gender-switching or for advantageous male reproductive function (which could also explain the lack of gender-switching).

Finally, "masculinization" is not only important at species level, but it is also a necessary assumption to explain phylogenetic patterns of mitochondrial DNA of DUI different species, genera or families (Hoeh et al., 1997) (see below).

### *2.4.4 DUI DISTRIBUTION AND PHYLOGENY OF PATERNAL AND MATERNAL GENOMES.*

As above mentioned, 36 species possessing DUI belong to seven families and five superfamilies all within Autolamellibranchiata. Within Autolamellibranchiata, DUI was found in Pteriomorphia as well as in Heteroconchia, and within Heteroconchia in Palaeoheterodonta and Heterodonta (Theologidis et al., 2008).

Because of its wide taxonomic distribution and its scattered occurrence, there has been some discussion as to whether DUI evolved one or many times among Bivalvia. However, because of the evident analogies of DUI in all the analyzed species, the most sound hypothesis is that DUI originated once at the evolutionary radiation of Bivalvia.

If we assume that DUI arose only once, its broad taxonomic distribution makes DUI a very old phenomenon within the animal kingdom, perhaps as old as the origin of bivalves, that is 506 million year ago (Mya), but definitely as old as the origin of Autolamellibranchiata, 460 Mya (Little and Vrijenhoek 2003). Moreover, we need to admit that the reversions to SMI have occurred many times, thus producing the scattered distribution of DUI known so far (Breton et al., 2007 and references therein).

A particular caution has to be put in assessing the absence of DUI in a bivalve species, since most of the time the invention of mtDNA homoplasmy has been considered as an evidence for SMI: this is not necessarily true, since in DUI heteroplasmy is achieved only if M and F mtDNAs never switch their roles over a certain amount of time, in which genomes can accumulate mutations. However, we know that DUI mtDNAs can switch their roles, and especially F mtDNA can invade male line and be transmitted through sperm, the event known as "masculinization" (see above). This will re-set to zero the divergence among genomes. Actually, DUI might work quite well even if role-reversals of sex linked mtDNAs occur frequently.

If we look at taxa with DUI, the gender-linked mitochondrial genomes could cluster in three alternative ways: a "gender-specific" pattern if the two female genomes do form one cluster and the two male genomes another (*e.g.* Unionidae); a "taxon-specific" pattern, if female and male genomes from each taxon do cluster together (*e.g.* Veneridae); finally a "mixed pattern" if the clustering pattern result to be may be random regardless of taxon and gender (*e.g.* Mytilidae).

Theologidis et al. (2008) constructed a phylogeny of the maternal and paternal genomes of DUI species, to address the question of a single or multiple origins of DUI. They could observe that F and M sequences from the same family branch together away from the F and M sequences from other families. An exception to this comes from Unionoidea, where F and M sequences from two families, Unionidae and Hyriidae, cluster according to the gender-specific pattern, an observation also made previously by Hoeh et al. (1996, 1997, 2002). In this superfamily the F and M phylads have diverged to the point that when grouped with sequences from other families their common origin is weakly supported or not detectable.

The authors offered two possible explanations to this: the simplest would require one DUI origin for all Unionoidea, and one for each of the families Solenidae, Veneridae, Donacidae, and Mytilidae. Moreover in addition to these, a separate origin would be required for each of the four mytilid genera and the two genera of Veneridae that, at present, are known to have DUI. This hypothesis would need to postulate a larger number of independent DUI origins, as the search for DUI in other bivalve species continues. An additional difficulty with the hypothesis of multiple origins is the molecular and developmental complexity that likely underlies the phenomenon of DUI and its

linkage to sex inheritance (Zouros 2000; Cogswell et al. 2006). The repeated and independent emergence of this complex mechanism would imply a strong selective advantage of DUI over the standard maternal inheritance.

By contrast, the alternative explanation of a single origin of DUI cannot by itself explain the phylogenetic relationships among the different taxa, if it is not combined with masculinization. Seen in phylogenetic context, the hypothesis consists of two phases, the emergence of a new sperm-transmitted genome from an egg-transmitted one and the replacement of the previously sperm-transmitted genome in the population by this new genome. Moreover they supposed that the time lapsed since the separation of unionids, venerids, solenids, donacids, and mytilids was enough for at least two masculinization events to have occurred in the lineage leading to each of these groups. As discussed by Hoeh et al. (1997), masculinization events have apparently not occurred after the emergence of unionids, either because transmission reversals have not occurred with any substantial frequency or because newly masculinized genomes could not spread in the population. The possibility that there has been not enough time for masculinization to complete its course is unlikely given the large deviation of the unionid F and M genomes.

### *2.4.5 THE MYTILID MUSCULISTA SENHOUSIA: AN UNUSUAL CASE OF GENDER-ASSOCIATED MITOCHONDRIAL DNA HETEROPLASMY.*

Passamonti (2007) gave first evidence for a new DUI case of mitochondrial DNA in the mytilid *Musculista senhousia*. The heteroplasmic pattern obtained was in line with other DUI systems of mtDNA inheritance: sperm and eggs carry two different mitochondrial haplotypes, M and F respectively, so the author assumed that two different mitochondrial genomes are passed to progeny. Although tissue distribution of female and male haplotypes in *Musculista* appeared to be in line with other mytilid species, M haplotypes were relatively rare in the male soma. This kind of distribution allowed to speculate on the mechanisms transmitting M mitochondria to progeny. The egg carries many more mitochondria than the spermatozoon; moreover it has also been observed that sperm mitochondria enter the egg during fertilization (Longo and Dornfield, 1967; author's personal observations on *Tapes philippinarum*). According to this, it isn't a simple conjecture to postulate that, in the zygote, F mitochondria (which derived from the egg), must overwhelm the M ones, inherited from the spermatozoon.

According to observations of both Cao et al. (2004) and Cosgwell et al. (2006) (see below), male mitochondria have to find their way to blastomer 4d, which is known to give rise to germ line (Verdonk and Van Den Biggelaar, 1983). This may happen in two ways: the first is that M type

should undergo several replications during the early male embryo cleavage, so that M mitochondria will have a good chance to be included in the 4d blastomer (Sutherland et al., 1998); the second, is that sperm mitochondria are actively segregated to the 4d blastomer itself. In both cases, the M mitochondria displace F ones during testis development and, if the first scenario is true, then we may expect a large quantity of M mitochondria in the male somatic tissues of adults, while, if the second is true, we may expect no or few M genomes in the soma. It was a reasonable opinion of author that the second scenario seemed to fit better the case of *Musculista* because M haplotypes were extremely rare in male somatic tissues of this species. Thus, he assumed that M mitochondria are actively segregated to germ line. In fact, this pattern is in line with the recent observations of the cytological behavior of sperm mitochondria in early embryo stages. These latter observations showed that the second scenario seems also true for *Mytilus edulis*:, it was shown that sperm-derived mitochondria (using in vivo mitochondrial staining) are actively segregated to blastomer 4d in the early cleavages of male embryos (Cao et al., 2004). It must be noted, however, that M mitochondrial DNAs are also detectable in specific *Mytilus* somatic tissues (this is also confirmed in *Musculista*), so that the M segregation to blastomer 4d must not be a highly selective mechanism.
Although the discovery of sex-related heteroplasmy has not been surprising for a mytilid, one trait of *Musculista* DUI system were seemed to be unique, contrasting with all previously described DUI systems above reported: the analysis evidenced that F haplotypes sequence variability was constantly higher than that of M haplotypes. Such inverted variability pattern (the F being more variable than M) was new and unexpected, and it challenged most of the rationales proposed to account for sex-linked mtDNA evolution. Moreover, the author also evidenced that *Musculista* F mtDNAs showed a higher mutation level in male's soma, this variability being produced de novo each generation; this was in line to Borras et al. (2003) that found that mitochondria from females exhibit higher antioxidant gene expression and lower oxidative damage than males. This might have an evolutionary explanation: mechanisms evolved to protect mtDNA in females (f.i. antioxidant gene complexes) might be under relaxed selection in males, because males are evolutionary dead-ends for mitochondria. This is true both for SMI species and DUI species, since female-transmitted mitochondria are not passing to progeny through males in both cases.
These results suggest that higher variability could be a general feature of the whole *Musculista senhousia* mitochondrial genome, but the ultimate demonstration will be available only with the complete M and genomes are sequenced. For this reason I proceeded to the complete sequencing of both sex-linked mitochondrial genomes of *Musculista senhousia* and the result has been surprisingly (see Results and Discussion).

## *2.5 THE MITOCHONDRIAL GENOME IN INSECTA*

As reported above, mitochondrial genomes are one of the most information-rich molecular markers in phylogenetics, because of more than 15,000 bp of nucleotide data and 37 genes. In addition to the nucleotide data, other phylogenetically useful information can be extracted from the mt genomes, such as gene rearrangements (Boore and Brown, 1998), gene insertion or deletion events (Rokas and Holland, 2000), and genic or intergenic length variability (Schneider and Ebert, 2004). Gene rearrangements have been used to generate phylogenetic trees since 1930's (Dobzhansky 1944). It is usually assumed that rearrangements are rare because of the requirement for two or three chromosomal breaks, as it seems unlikely that identical chromosome breaks would occur in independent lineages, and that such breaks are reversible.

Since mitochondrial genome provides both the largest set of homologous genes which can be compared across animals, and the largest stretch of molecular data which can be readily used in comparisons of gene order, these features would appear to make mtDNA rearrangements ideal cladistic markers.

Two approaches to the phylogenetic use of the mitochondrial genome have been proposed, *i.e.* shared genome rearrangements and whole genome sequence based phylogenies (Boore and Brown, 1998). Some of the first insect mitochondrial genomes to be sequenced, *Apis mellifera* (Hymenoptera) (Crozier and Crozier, 1993), *Locusta migratoria* (Orthoptera) (Flook et al., 1995a), *Heterodoxus macropus* (Phthiraptera) (Shao et al., 2001) have moderate gene order rearrangements relative to the arthropod round-plan suggesting that the use of "genome morphology" (Dowton et al., 2002) would be a viable approach to resolving deep nodes in insect evolution. The subsequent sequencing effort has not supported this promise as useful phylogenetic marker for the resolution of interordinal relationship, thus leading to an increasing emphasis on using the mitochondrial genome sequences in phylogenetic reconstruction.

In fact, most insect orders share the same plesiomorphic genome arrangement (Cameron et al., 2006) that is shared by the Pancrustacea (Boore et al.1998), and those orders that differ are therefore autapomorphic. In this regard, genomes rearrangements in Insecta are mostly related to translocations, losses or acquisition of *tRNA* genes, involving above all those adjacent to the control region or those surrounding a *tRNA* gene cluster [*trnA- trnR- trnS(AGN)- trnE- trnF*], which is considered as the replication origin of the mitochondrial genomes minor strand (Boore, 1999)

For what above reported, it appears clear that gene rearrangement synapomorphies will not contribute too much to resolve insect relationships. While the possibility to find additional rearrangements within insects' orders does exist, it is doubtful that such findings would greatly aid the understanding of inter-Order relationships. Moreover, in most cases, taxon sampling was

focused on capturing the diversity within each order, such as by sequencing representatives of each suborder or major clades, or including earliest branching representatives: the possibility that those species actually represent secondary reversions to the insect ground-plan is therefore low and it is doubtful that additional interordinal synapomorphies will come to the light. In contrast, the potential for using genome rearrangements in understanding intra-Order relationships seems good. Rearrangements have been found in over a third of the insect orders and in those orders where multiple representatives have been examined the phylogenetic signal in rearrangements is very strong.

Differently from genome rearrangements, sequence-based phylogenies have ranged from resolving strains within *Drosophila simulans* (Ballard 2000a) to relations among arthropods (Nardi et al., 2003a; Cameron et al., 2004) providing some remarkable results such as inferring Phthiraptera+Hymenoptera (Nardi et al., 2003a), the polyphyly Hexapoda (Nardi et al., 2003a; Bae et al., 2004) and monophyly of Orthoptera (Cameron et al., 2008).

The phylogenetic utility of mt genomes has been studied rigorously in the past years, especially for insects and related groups (Cameron et al., 2007, 2004; Carapelli et al., 2007; Kjer and Honeycutt, 2007). These results have highlighted the need for rigorously evaluating the phylogenetic behavior of mitochondrial genomes to improve confidence that results are reflective of evolutionary history rather than analytical artifacts that do not reflect the true phylogeny.

Whilst the capacity to quickly sequence insect mitochondrial genome is now a routine, questions remains to how best to analyze these data.

Actually, phylogenetic studies of insects, especially that addressing relationships among higher taxa, often exhibit a portion with low support or resolution. When a large amount of data has been analyzed, a possible cause for the poor support is rapid radiation, because it would result in truly short divergence times for diagnostic mutations to occur. Unfortunately, such a signature of a rapid radiation can also result from, or alternatively be obscured by, a variety of other causes related to data quality. It is therefore important to test whether the available data are appropriate to resolve relationships at the hierarchical level being analyzed, and to determine whether confounding biases in the data are interfering with phylogenetic signal detection.

Actually, what is an “ancient rapid radiation”? It’s a case of diversification in which lineages of insects have diverged in rapid succession within a relatively short time span in the ancient past, generating patterns of molecular and morphological change that are difficult to discern phylogenetically.

Although insects diverged spectacularly in the Permian, and again in the Jurassic, and diverged ever since, “ancient” does not necessarily refers to a specific age, but rather to a high ratio between the

time since divergence occurred and the time span in which it occurred. Such high ratio is of course higher for cladogenetic events that are many millions of years old.

Rokas et al. (2005) described the typical signatures of molecular phylogenetics that are typical of an ancient rapid radiation: the signature can be described in a phylogeny as a significantly closer temporal spacing (compression) of a number of lineage-splitting events than would be expected by either stochastic or relatively constant diversification. Rapid radiation signature can be obscured by other factors such as inadequate data, conflicts within or among datasets, or loss of phylogenetic signal over time. Unfortunately, many insect phylogenies are unstable because of short, ancient internodes, lineage-specific substitution rate biases (substitutions accumulate at different rates among lineages), and lineage-specific base compositional biases (the nucleotides are sometimes found in different proportions among lineages, and among regions of genes). These problems, sometimes combined with data which substitution rates are wildly inappropriate for the addressed questions, make insect phylogenetics a particularly challenging goal.

Multiple substitutions at the same sites may cause problems for deeper-level phylogeny, and make phylogeny estimation impossible for deeper nodes (Ho and Jermiin, 2004; Mossel and Steel, 2004; Mossel and Steel, 2005; Philippe and Adouette, 1996; Philippe and Laurent, 1998). A number of other biases can increase loss of signal, especially when they are non random: among them, variations in the nucleotide (or amino acid) composition among taxa, differences in rate of change among sites within genes, differences in the distribution of variable sites among taxa, and differences in the rate of sequence evolution among lineages, could be taken into account. Actually, base composition heterogeneity (*i.e.* significant differences in nucleotide proportions in homologous sequences) has been suggested to cause systematic wrong phylogenies in a number of studies (Collins et al, 2004; Gruber et al., 2007; Jermiin et al., 2004; Phillips et al., 2004). The difficulty comes from the fact that lineages sharing similar base-compositions tend to group together, even when they are distantly related. Differences in substitution rate among sites (ASRV: among-site rate variation, alternatively called rates across sites) are also important to take into account in phylogenetic analysis (Gaut and Lewis, 1995; Jin and Nei, 1990; Kuhner and Felsenstein, 1994; Tateno et al., 1994; Yang, 1995; Yang, 1996). The most broadly employed methods for taking ASRV into account in evolutionary models include incorporating an estimate of the proportion of invariable sites (Lockhart et al., 1996; Steel et al., 2000) and/or using a gamma distribution to model the distribution of rate variation (Yang, 1995; Yang, 1996). Simplified, the gamma distribution can fit the data to a curve in which many sites are expected to change slowly, and a few are expected to change many times. Another approach is to use site-specific rate models (Kjer and Honeycutt, 2007). Whatever the approach, most of the common computer-implemented

methods currently have the disadvantage of assuming that all the lineages analyzed show the same patterns of variation among sites, which may not be true.

If the lineages one is analyzing differ in their rates of change, and even in the patterns of nucleotide sites that are changing, the effect on analysis can be profound (Inagaki et al., 2004; Lockhart et al., 2006). The erroneous grouping of more distantly related lineages is often caused by such rate and pattern differences (Lockhart and Steel, 2005), especially when some lineages are characterized by long branches (the long-branch attraction problem: *i.e.* the tendency of phylogenetic methods to group long branches).

The above mentioned problems may apply to a particular gene and can be expected to be more significant in older divergences, and with faster-evolving (possibly substitutionally saturated) genes. It is better to use the slowest-evolving genes to analyze the oldest divergences; the problem for ancient divergences (especially 200–300 mya and older) is that no known genes may actually have the desired rate of variation.

Different methods of analyzing mt genomes have been proposed and applied: selecting a few key genes from the entire mt genome (Nardi et al., 2003), including all available genes but not the control region (Castro and Dowton, 2007), analyzing amino acid sequences (Nardi et al., 2001) or recoding of nucleotide sequences purine/pyrimidine (i.e. A+G as purine: R and C+T as pyrimidine:Y) (Saitoh et al., 2006), and traditionally analyzing nucleotide sequences (Cameron et al., 2007). Different optimality criteria and dataset compilation techniques have also been applied to find the best method of analyzing complex mt genome data (Cameron et al., 2004; Castro and Dowton, 2005; Kim et al., 2005; Stewart and Beckenbach, 2003).

According to Cameron et al. (2007), mitochondrial genome data recover the most phylogenetic signal when all available genes are analyzed as nucleotide sequences, and results from different optimality criteria are then compared for the sensitivity of resulting phylogenetic tree building methods.

Mt genome data were not able to unambiguously resolve the relationships among major arthropod lineages (Cameron et al., 2004), but were able to resolve intraordinal relationships within Diptera (Cameron et al., 2007) and Hymenoptera (Castro and Dowton, 2007). This suggests that mitochondrial genomes data may not have sufficient phylogenetic signal to resolve ancient Cambrian to Devonian cladogenesis (600–360 MYA), but may provide strong signals for resolving intra-Order relationships within Diptera and Hymenoptera whose first fossils are found dating to the Upper Triassic (225 MYA). These findings imply that the maximum resolving capacity of insect mitochondrial genomes may lie somewhere between these two geological times.

### *2.5.1 THE PALEOZOIC DIVERSIFICATION OF THE INSECT ORDERS AND THE "ORTHOPTEROIDS PROBLEM"*

Being the most diverse group of terrestrial animals, insects may supply many examples of ancient rapid radiations. Many of the most spectacular radiations involve interactions with angiosperm plants, which also radiated extensively during the Cretaceous and Tertiary. However, insects were already highly diverse by the Permian, long before the origin of angiosperms. In some cases we can only speculate why the insect groups diversified as rapidly as they did.

Some phylogenetic relationships among insect Orders are well corroborated and others are best described as a polytomy (*i.e.* a node in a phylogenetic tree that subtends more than two branches). Before the advent of molecular systematics, Kristensen (1975, 1981, 1991, 1995 and 1997) proposed a tree derived from morphological characters that contained Hexapoda, Insecta, Dicondylia, Neoptera, Dictyoptera, Paraneoptera, and Holometabola, but one major part of the tree was left unresolved, *i.e.* the relationships among the orthopteroid orders (or Polyneoptera).

Since then, molecular data have done little to gain a common view of insect Order relationships, with each study contradicting previous work (Chalwatzis et al., 1996; Liu and Beckenbach, 1992; Terry and Whiting, 2005; Wheeler et al., 2001; Whiting, 2002; Whiting et al., 1997) and virtually no congruence among independent datasets have been found, except for nodes that have been firmly established by morphology (Kjer et al., 2006).

Ten neopteran lineages are clustered together to form orthopteroid (or Polyneoptera) infraclass: Plecoptera, Dermaptera, Embioptera, Phasmatodea, Mantophasmatodea, Grylloblattodea, Dictyoptera (i.e. Blattaria, Isoptera and Mantodea), Zoraptera, Orthoptera, and Paraneoptera plus Holometabola.

The upper and lower time limits these 10 lineages diverged is within the first land plants evolution at 475–425 mya (Wellmann et al., 2003) and the extant fossil orders 280 mya ago: for the upper limit, it is hard to imagine terrestrial animal life of any kind before there were plants to feed upon and offer shelter; at the lower limit, fossil Plecoptera, Orthoptera, and Dictyoptera have been found in the Permian. Thus, all these lineages must evolved at least within 200 million years, although more likely their divergences are bounded by the emergence of the putative stem neopterans (Paoliidae) in the mid-Carboniferous (Grimaldi and Engel, 2005), which would mean that the orthopteroids diversified within less than 50 million years. That leaves a series of short cladogenetic events between 5 and 20 million years, with lineages that are over 300 million years old now. Such branches are therefore very long.

In this regard, the emerging phylogenetical scenario provided by the ten neopteran lineages is a "virtual polytomy". There is some molecular support for Plecoptera plus Dermaptera, Embioptera

plus Phasmatodea, and Grylloblattodea plus Mantophasmatodea (Kjer et al., 2006; Terry and Whiting, 2005). Two of these are contradicted by the mitochondrial data (Cameron et al., 2006, Kjer et al., 2006), which place Mantophasmatodea with Phasmatodea.

I have sequenced the nearly complete mitochondrial genomes of two *Bacillus* specimens (*Bacillus atticus* and *Bacillus rossius*) with the aim to better resolve phylogenetic position of Phasmatodea within orthopteroids lineages. At the time, mitochondrial genome of *Timema californicum* (suborder Timematodea) is the only available for Phasmatodea. However, *Timema* is the earliest diverging stick insect (Whiting et al., 2003), so that may not be really representative of the Order. The two *Bacillus* mitochondrial genomes here reported add to the phasmatodean taxon sampling, and are the first representatives from the Verophasmatodea suborder.

## 3 RESULTS AND DISCUSSION

### *3.1 MOLLUSCA BIVALVIA*

The mitochondrial genome sequences of *Musculista senhousia* here obtained are 20,612 bp in male (M-type or M) and 21,557 bp in female (F-type or F). The size of both M and F mitochondrial genomes are well within the size range of mollusk mtDNAs sequenced to date, *i.e.* from 13,670 bp in *Biomphalaria glabrata* to 32,115 bp in *Placopecten magellanicus*. *M. senhousia* gene arrangement is remarkably different from that known for other fully sequenced metazoan mtDNAs (Gissi et al., 2008 for a review) (Fig.1; Fig.2 and Tab.1). Both M and F mitochondrial genomes include a large number of unassigned regions (27 and 29 in male and female respectively). Among these, the largest (2,847 and 4,567 bp in male and female respectively) are apparently the main CR (here referred as LUR, i.e. Large Unassigned Region) for replication and transcription.

#### *3.1.1 OVERALL GENE ORDER*

A remarkable observation is that M and F mt genomes showed a different gene order. Female mt genome of *M. senhousia* contains the full complement of genes of the typical metazoan mtDNA, but lacks the ATPase8 subunit gene, while an extra *trnM* and *trnL* gene are present. Male mtDNA retains the same female's gene content except for the *cox2* gene which is duplicated in the M genome only and for the *trnE* gene which is traslocated (Fig.1and Fig.2; Tab.2 and Tab.3).

In both genomes, all genes are transcribed from the same strand except for *trnS*, which is transcribed in the opposite one. Most of the marine bivalves whose full mtDNA sequence is known to date (*Mytilus* species-complex, *Crassostrea gigas*, *C. virginica* and *C. hongkongensis*; *Venerupis philippinarum*) share the common feature of all genes coded by the same DNA strand. In contrast,

this is not true in the two freshwater species *Lampsilis ornata* (Serb and Lydeard 2003) and *Inversidens japanensis*. In other mollusks, a relatively small number of mitochondrial genes are transcribed from the second strand. The scaphopods *Graptacme eborea* and *Siphonodentalium lobatum* are an exception, with about an equal number of genes encoded by each strand (Boore, Medina, and Rosenberg 2004; Dreyer and Steiner 2004). The occurrence of all genes in the same strand is a relatively rare phenomenon in metazoans and, in addition to bivalves, it has been reported in some annelids (*Lumbricus terrestris*, Boore and Brown 1995*; Platynereis dumerilii*, Boore and Brown 2000) and brachiopods (*Terebratulina retusa*, Stechmann and Schlegel 1999; *Terebratalia transversa*, Helfenbein, Brown, and Boore 2001; *Laqueus rubellus*, Noguchi et al. 2000).

### *3.1.2 NUCLEOTIDE COMPOSITION*

The nucleotide compositions of the two genomes are summarized in Table1. The G+T content of the F and M coding strand is 60.7% and 59.6%, respectively, and thus the sense strand can be considered as the heavy (H) strand of the molecule. The A+T content of the H strand is also high (66.5%, F; 67.0%, M). Variable values of A+T content are common in mollusks, and they have been reported in *Lampsilis ornata* (62%, Serb and Lydeard 2003), *Pupa strigosa* (61.1%, Kurabayashi and Ueshima 2000), and *Cepaea nemoralis* (59.8%, Yamazaki et al. 1997). In other mollusks, the A+T content is much higher *(Albinaria coerulea*, 70.7%, Hatzoglou, Rodakis, and Lecanidou 1995; *Katharina tunicata*, 69.0%, Boore and Brown 1994; *Graptame eborea*, 74.1%, Boore, Medina, and Rosenberg 2004). This variation in A+T content is among the highest observed within a phylum and reflects the high heterogeneity of molluscan mtDNA (Boore 1999). Moreover, there is a marked bias in favor of T against C, which is not restricted to any particular class of genes and does not differ between the two genomes. The GC and AT asymmetry between the two mitochondrial DNA strands can be expressed in terms of GC skew and AT skew calculated according to Perna and Kocher (1995): GC skew = (G-C)/(G+C) and AT skew = (A-T)/(A+T), where G, C, A, and T are the occurrences of the four bases in the H strand. In *M. senhousia* F and M mitochondrial genomes, the GC skew and the AT skew are F: +0.28 and -0.18, and M: +0.23 and -0.17, respectively.

### *3.1.3 RIBOSOMAL AND TRANSFER RNA GENES*

The location of the two ribosomal RNA genes, obtained through BLAST comparison, do not differ between male and female mitochondrial genomes. However, their length is different between the two mitotypes. More in detail, in both sexes, *rrnaL* is located in a region flanked by the *trnM2* and *nad3* genes; assuming that the first base at the 5'-end come immediately after the *trnM2*, and the 3'-end of the gene correspond to the first base upstream the start codon of *nad3* gene, the length is remarkably different: the male *rrnl* gene (1,682 bp in length) is 557 bp longer than the female one (1125 bp in length).

The *rrnS* genes is located in a region flanked by *trnS* and *nad6* genes, and as above, we assumed that the first base at the 5'-end come immediately after *trnS*, and the 3'-end of the gene correspond to the first base upstream the start codon of *nad6* gene. Here, the difference in length is less remarkable between the different sex-mitotypes: actually, the female *rrnS* gene (819 bp long) is shorter 268 bp than the male one (1,087 bp long).

The identification of tRNA genes was based on their potential to form cloverleaf structures. F and M genomes of *M. senhousia* contain 22 tRNA genes, one more than is typical for the metazoan mtDNA. As well as in mtDNA of some other mollusks (*Katharina tunicata*, *Cepaea nemoralis*, *M. edulis* spp. and *Argopecten irradians*), two leucine tRNA genes are present in *M. senhousia*, which can be differentiated by their anticodons (TAA for *Leu1*, TAG for *Leu2*). Similarly, an additional *trnM*, was also detected, as found in female-*V. philippinarum*, *Mytilus species*, complex, in *Crassostrea gigas*, *C. hongkongensis* and *C. virginica*. The additional tRNA coding for methionine has the TAT anticodon.

In both male and female, *TrnS* have shortened DHU arm stems, where potential pairing consist of only one base pairs in mispairing (A-A.). This is not atypical, as the DHU arm of the *trnS* (AGN) is unpaired in many metazoan taxa (Garey and Wolstenholme 1989; Hoffman, Boore, and Brown 1992; Yamazaki et al. 1997; Tomita et al. 2002). Moreover, mispairing between bases in stems is consistent across several taxa. For example, the second base pair in the anticodon stem of *trnW* has a T-T mispairing in *L. ornata*, *Mytilus*, and *K.tunicata* and a T-G pairing in several gastropods (Yamazaki et al. 1997).

### *3.1.4 GENE JUNCTIONS*

Ojala, Montoya, and Attardi (1981) suggested that the secondary structure of a tRNA gene between a pair of protein genes is responsible for the precise cleavage of the polycistronic primary transcript.

In the absence of an intervening tRNA, this role can be played by a stem-loop structure, the 5'-end part of the gene itself, or a combination of the two. Potential hairpin structures at protein-protein gene junctions with no intervening tRNA have been reported in several studies (e.g., Bibb et al. 1981; Clary and Wolstenholme 1985; Okimoto et al. 1992; Boore and Brown 1994; Mizi et al., 2005).

In the F mtDNA of *Musculista*, 20 out of 22 tRNA genes are clustered in five groups of two to six (see Tab.3). Of the remaining two, *trnT* lies between *atp6* and the 5'-end of *cob* genes (with 24 bp overlapping each other) while *trnA* lies between *nad5* and *nad4* genes. Thus, 4 of the 12 protein-coding genes (*cob*, *nad1*, *nad4* and *nad4L*) have a tRNA preceding their 5'-end. On the contrary, five other genes (*cox1*, *cox2*, *atp6*, *nad2*, *cox3*) have a noncoding sequence preceding their 5'-end that is capable to form a stem and loop structure (Fig.3). This seems to be not true for the *nad5* gene. In all these genes, a putative stem-and-loop structure that includes the translation initiation codon can be formed downstream from the 5'-end part of the gene.

In male mitochondrial DNA, 19 of the 22 tRNA genes are clustered in five groups of two to six members (see Tab.2). Of the remaining three, *trnT* lies between *atp6* and the 5'-end of *cob* genes (with 24 bp overlapping as observed for the female), *trnA* lies between *nad5* and *nad4* genes and *trnE* lies between the large unassigned region (LUR) and the 5'-end of *cox1* gene. Thus, 5 of the 12 protein-coding genes (*cox1*, *cob*, *nad1*, *nad4L* and *nad4*) have a tRNA preceding their 5'-end.

Five other genes (*cox2a, cox2b, atp6, nad2, cox3*) have a noncoding sequence preceding their 5'-end that is capable of forming a stem and loop structure (Fig.3). As above, this seems to be not true for the *nad5* gene. In these five genes, a putative stem-and-loop structure that includes the translation initiation codon can be formed downstream from the 5'-end part of the gene.

### *3.1.5 GENE CONTENT*

The protein-encoding genes are of the same length in the two genomes, except for *nad1*, *nad3*, *nad4* and *nad5*, in which the difference is 1, 5, 3 and 5 amino acids, respectively (Tab.4).

As mentioned, the male *Musculista senhousia* mitochondrial genome contains an extra copy of the *cox2* gene. This is not new for DUI bivalves, since also the female mt genome of the marine clam *Venerupis philippinarum* contains the same duplication.

In the female *Musculista*, the *cox2* gene (here referred as F-*cox2*) is 660 bp long and is flanked by the "*cox1*/UR-6" and "UR-7/*atp6"* regions at the 5'- and 3'-end respectively (see Tab.1 and Fig.4).

In male mitochondrial genome, the two copies of *cox2* are close to each other and linked by a little non coding region 41 bp long (here referred as UR-6). The two *cox2* copies are located between "*cox1*UR-5" and "UR-7/*atp6*" regions, and the first is 813 bp long, while the second is 690 bp long.

Analyses on the three sequences evidenced that F-*cox2* is more closely related to the shorter M-*cox2* (690 bp), rather than to the longer M one; therefore, the 813 bp long M-*cox2* seems to be an extra copy of the gene, and thus it will be referred here as M-*cox2b*.

In order to analyze the structure and evolutionary forces shaping the duplicated *cox2* gene in the male genome, I started comparing their nucleotide and aminoacid sequences. Nucleotide P distance values do not provide any particular indication, because all comparisons (*i.e.* F-*cox2 vs.* M-*cox2* and F-*cox2 vs.* M-*cox2b*) evidence quite similar values (0.264±0.016 and 0.267±0.016 respectively). In contrast, synonymous (Ks) and non-synonymous (Ka) substitution values are interesting and enlighten a quite different scenario: in fact, the comparison between F-*cox2* and M-*cox2* gave a Ks of 0.887±0.124, Ka=0.178±0.024 and a Ka/Ks ratio of 0.20, which are quite in line with the average values obtained by comparing all the other M and F protein coding genes (0.750±0.022; 0.146±0.005 and 0.19 respectively). This result might be taken as an indication, although preliminary, that both genes are evolving in a neutral way. The scenario radically changes comparing Ks and Ka values between the two male's *cox2* genes: non-synonymous substitutions are now more common (Ka=0.223±0.030), while synonymous are less (Ks=0.653±0.016), and this obviously reflects in a higher Ka/Ks ratio (0.34). This scenario would seem to suggest that the M-*cox2b* gene may evolved under positive selection.

Moreover, a different pattern of amino acid replacements is evident when comparing F-*cox2n vs.* M-*cox2* and M-*cox2b vs.* M-*cox2*, respectively: in the first, 5 out the 55 aminoacid substitutions are from hydrophilic to hydrophobic amino acids and vice versa, *i.e.* only 10% of the substitutions. In the second comparison, this percentage increases to 25%. This also seems to provide a further indication that the duplicated *cox2* gene (M-*cox2b*) might be somehow selected for a different function in males. A more detailed analysis on protein structure, not possible at the moment, is planned to address this.

Stronger evidences are still lacking, and this is a characteristic that will be investigated further, but some hints may come from an interesting analogy to unionid mtDNA: actually, in freshwater clams, the *cox2* gene seemed to be somehow involved in the DUI mechanism itself, and a COX2 protein function in tagging the male mitochondrial genome has been proposed, since the authors evidenced that the male COX2 is present on both inner and outer membranes of the mitochondrion (Chakrabarti et al., 2006). At variance to *Musculista*, unionids *cox2* putative mechanism was not related to a gene duplication, but to a peculiar 3' extension of the male *cox2* gene (Curole and Kocher, 2002, 2005; Chakrabarti et al., 2006). It is conceivable however, that *Musculista* and Unionid unusual *cox2* features are somehow related, *i.e.* they might share a common function: actually, the duplicated *cox2b* in male *M. senhousia* mitochondrial genome may represent just a

variant of that found in unionoidean bivalves, with proper signals for DUI mitochondrial tagging lying in the *cox2* extension of unionid bivalves, and in the *cox2b* duplicated gene of *Musculista*. Finally, it is also not possible to exclude that control mechanism of developmental signals are affected by redundancy.

Moreover, no *atp8* coding sequence was detected. Boore (Boore,1999) mentions that a lack of the *atp8* gene is one of several unusual features of the *Mytilus* mt sequence. The *apt8* gene is missing from the mt-DNA of almost all bivalve species studied so far, including *C. hongkongensis*, *C. gigas*, *C. virginica*, *Mytilus*, *P. magellanicus*, *A. irradians*, *M. yessoensis* and *Acanthocardia tuberculata*. The one exception found so far is *Hiatella arctica* in which the *apt8* gene is present. Interestingly, *atp8* gene is present in the female mitochondrial genome of the unionid bivalve *L. ornata* (Serb and Lydeard, 2003). A remarkable observation is that *V. philippinarum* was recently found to contain a putative *atp8* gene (Dreyer and Steiner, 2006), which was not found in the first analysis. Nonetheless, this gene apparently encodes 37 amino acids only and therefore has a questionable gene function. In contrast, all Gastropoda species (Yamazaki et al., 1997) studied so far do possess an *atp8* gene, as they do all Cephalopoda (Rawlings et al., 2001). Other mollusk species, from Polyplacophora *K. tunicata* to Scaphopoda *G. eborea* and *Siphonodentalium lobatum*, have an *atp8* gene as well.

### *3.1.6 COMPARATIVE SEQUENCE ANALYSIS*

Comparing the two sex linked genomes, it appears that the most conserved protein-coding genes in *Musculista* are *cox1* and *cob* (identity > 99%), and the least conserved are *nad2* and *nad6* (identity < 70%, (Tab.4).

Synonymous (Ks) and non-synonymous (Ka) substitution values between the two genomes vary (Tab.1 and Tab.4). Ka is particularly low for *cox1* (0.042±0.006), but Ks is not (0.838±0.073), suggesting that this gene is under some selective constraint (Ka/Ks=0.05). The conservation of *cox1* is common in animal mtDNA (Pesole et al. 1999; Saccone et al. 1999). In *cob* gene, both K values are lower than average (table aa) with a Ka/Ks ratio's value (0.09) which is close to that of *cox1* gene. In contrast Ks and Ka deviate from average in *nad6*, but in opposite ways, thus providing the most value of Ka/Ks ratio (0.43) among those observed (Tab.1 and Tab.4).

Mizi et al., (2005) found a similar situation in the *Mytilus edulis* species complex, but with an interesting exception: they realized a comparison of both synonymous (Ks) and non-synonymous (Ka) values between *Mytilus'* male and female genomes and observed the same pattern for *cox1* and *nad6* genes as here reported for *Musculista*, but, in contrast, both k values were lower than average in *nad3* gene, which is different from what observed here for *Musculista senhousia* (for *cob* gene).

Nonetheless, they extended the comparison to the *V. philippinarum* and *I. japanensis* mitochondrial genomes. Again, in both pairs the *cox1* had the lowest divergence and the *nad6* divergence was among the largest. In contrast, in neither pair the *nad3* divergence was particularly low. Thus, they concluded that the low rate of evolution of *nad3* relative to other protein genes seems to be a characteristic property of the *M. edulis* species complex. As a possible explanation, they added that the conservation of the *nad3* gene in the *M. edulis* species complex might lie in the fact that *nad3* contained the origin of the replication of the lagging strand (L. Cao, E. Kenchington, A. Mizi, G. C. Rodakis, and E. Zouros, unpublished data in Mizi et al., 2005).

If this is true for *M. edulis nad3* gene, we may speculate that it also true for *cob* gene in *Musculista:* actually*, cob* gene exhibits the same identical pattern of k values (strongly lower than average, table aa) as well as that reported for *nad3* gene in *M. edulis* species complex by Mizi et al., (2005). As a consequence, we can speculate that the conservation of the *cob* gene in *Musculista* might be due to the fact that this gene may contain the origin of the replication of the lagging stand.

### *3.1.7 CODON USAGE*

In the *M. senhousia* male mtDNA six out 13 protein genes start with the ATA codon and seven with ATG, while in the female ATG and ATA start codon are equally distributed. This pattern is not in line to that observed for *Mytilus*, where nine out of the 12 protein genes start with the ATG codon and two with the ATA (Mizi et al. 2005). In all known metazoan mtDNAs, the most common start codon is ATG, and it is a general opinion that the methionine tRNA with the CAT anticodon represents the ancestral form. Moreover Hoffmann, Boore, and Brown (1992) suggested that the second methionine tRNA arose by duplication. The F and M genomes of the venerid *V. philippinarum* also have two tRNA genes for methionine, but both have the ‘‘ancestral’’ CAT anticodon. TAA codon is the termination codon nine and ten times, in female and male respectively, while only in two cases stop codon is TAG. In both sex, *Nad5* gene is terminated by an incomplete termination codon T--, with their likely completion occurring by polyadenylation after transcript processing (Fernandez-Silva et., 2003).

A total of 3834 and 3774 amino acids are encoded by male and female *Musculista senhousia* mitochondrial genome respectively. All codons do occur in both *Musculista* mitochondrial genomes (Tab.5). TTT (phenylalanine) is the most frequent codon, followed by TTA (leucine). TTT is also the most frequent codon in *Mytilus galloprivincialis* (Mizi et al., 2005), in *L. ornata* (Serb and Lydeard 2003) and in *C. nemoralis* (Terrett, Miles, and Thomas 1996), whereas TTA (leucine) is most common in *A. coerulea* (Hatzoglou, Rodakis, and Lecanidou 1995), *P. strigosa* (Kurabayashi and Ueshima 2000), *Roboastra europaea* (Grande et al. 2002), *G. eborea* (Boore, Medina, and

Rosenberg 2004), and *K. tunicata* (Boore and Brown 1994). These two codons are also the most frequently used codons in other invertebrate mtDNAs (Garesse 1988; Cantatore et al. 1989; Okimoto et al. 1992; Asakawa et al. 1995; De Giorgi et al. 1996; Helfenbein, Brown, and Boore 2001). TTT is also very frequent in primitive chordates (like amphioxus, *Branchiostoma lanceolatum*, Spruyt et al. 1998), but not in most vertebrates, where CTA (e.g., *Cyprinus*, Chang, Huang, and Lo 1994; *Homo sapiens*, Ingman et al. 2000) or ATT (e.g., *Xenopus laevis*, Roe et al. 1985; *Danio rerio*, Broughton, Milam, and Roe 2001) are the most frequent.

Serine can be specified by six different codons, *i.e.* UCY and AGN codons. However, our analysis with tRNASCAN-SE software failed to recover a potential cloverleaf structure for serine with AGN anticodon in both *Musculista* mitochondrial genomes. Conversely, as above mentioned, all codons for serine do occur in *Musculista senhousia* mitochondrial genomes and even more surprisingly, the AGN codon is more common than the UCY codons (261 and 115 times respectively). Further investigations are necessary to solve this issue.

The four least used codons are TCC (10 times in F), CGC (6 times in F), CCG (8 times in both F and M) and ACC (7 times in F). Of these, CGC is also among the least common in the mtDNA of other mollusks.

Synonymous codons, whether fourfold (4FD) or two-fold (2FD) degenerate, are recognized by the same tRNA (Tab.5), with the exception of the methionine codons, which are recognized by a different tRNA (Tab.5).

Moreover, 2788 *Musculista* codons (72.5% and 72.9% in female and in male respectively) end with an A or T, a more pronounced phenomenon than observed in the typical invertebrate codon bias. There is a strong bias against the use of C (9.6% and 10.9% in female and in male respectively) at the third position nucleotide in all codons: in detail, for amino acids with a fourfold degenerate third position, codon families ending with T are the most frequently used (50.3% and 46.9% in female and male respectively). This is also the case for twofold degenerate codons. In other words, in every case where an amino acid can be specified by any NNY codon, both female and male *Musculista senhousia* mitochondrial genome have a much higher proportion of NNT:NNC. In fact, female showed 49.1% of T and 12.5% of C, with NNT:NNC ratio of 4:1; while in male the ratio's value is a little inferior than the female one: 3.65:1 (45.6% of T and 12.5% of C). At the second position, there is even a stronger bias in favor of the use of T (46.2% and 45.2% in female and male respectively), which is also true for *M. edulis* (43.5%), *C. hongkongensis* (42.5%), *C.gigas* (42.3%) and *C. virginica* (43.0%).

Finally, in seven 2FD and eight 4FD families the most frequently used codon does not match the tRNA's anticodon. This has been observed in other metazoan mtDNA as well (Roe et al. 1985;

Wolstenholme 1992, Rand and Kann 1998; Crease 1999; Broughton, Milam, and Roe 2001) and suggests that strict codon-anticodon complementarity has not affected the codon composition of the genome.

Deviations from equal frequency of the four nucleotides in 4FD sites are common in the animal mtDNA and have been attributed to several factors, such as unequal presence of the four nucleotides in the nucleotide pool, preference of the mitochondrial gamma DNA polymerase for specific nucleotides, or asymmetrical mutation rate owing to different duration of exposure of the lagging strand during replication (Sueoka 1962; Asakawa et al. 1991; Jermiin and Crozier 1994; Jermiin et al. 1994, 1996). In Mizi et al., (2005) the author reported an unpublished data from Cao, Rodakis and Zouros: they have noted that 4FD sites with longer exposure have a higher probability to be a T and a lower probability to be a G. This observation implies that the frequency in which the four codons of a given amino acid are used depends on the distribution of the amino acid's residues along the genome. Yet, these correlations account for less than a third of the bias in codon usage and cannot explain the bias in 2FD codon families. Thus, other factors must also affect codon bias.

One among these may be the degree with which tRNAs recognize their different synonymous codons. If this degree varies among codons, the tRNA itself may act as a selection factor for synonymous mutations and may thus determine the frequency of synonymous codons in the genome.

### *3.1.8 LARGE UNASSIGNED REGION (LUR) AND SEX-LINKED MT-DNA TRANSMISSION*

In the male genome, the largest unassigned region is 2,847 bp long, and it is included between *trnN* and *trnE* genes. The female one is 4,567 bp long and it is included between *trnE* and 5'-end of *cox1* genes (Tab.1; Tab.6 and Tab.7).

Both male and female LUR start with an unrelated sequence/spacer 238 and 20 bp long, respectively. Male LUR is composed by a unique sequence (here referred as α), 1,752 bp long, followed by a shorter sequences (here referred as β), 267 long; this latter is tandemly repeated 4 times, the latter element being truncated (70 bp long) (Fig.5). In the first β repeat a palindromic motif has been found between nucleotide position 8-21. In each of the three following β repeats a larger one has been found, between nucleotide position 1-28. Both α and β sequences can also be found in the female LUR, but with a quite different organization (Fig.5). At variance of male-LUR, β repeat is not tandemly repeated, but is followed by a 160 bp sequence (here referred as γ) and the 2149 bp long block of α+β+γ is tandem duplicated. Downstream this duplication, a 202 bp fragment highly similar to the 5'-end of an α sequence is observed. Also, palindromic motifs are present in the female β repeats. While female and the first male β repeat show a 18 and 14 bp

palindrome motif, respectively, duplicated male β sequences (including the truncated one) show a longer motif (28 bp), therefore resulting in a longer stem structure (Fig.6).

In F LUR, pD-values obtained by comparing each pair of homologous repeats located in the two mayor repeats are 0.0023±0.0011, 0.0076±0.0052 and 0.0125±0.0085 for α, β and γ respectively. In M LUR pD-value for β repeats is 0.0027±0.0078. The values between sexes are 0.15±0.0091 and 0.34±0.026 for α and β repeats, respectively. On the whole, the pattern of variability evidences that the repeats have a higher sequence homogeneity within than between sexes. This would seem to suggest mechanisms of concerted evolution leading to homogenization of the repetitive region.

Moreover, we recovered a putative trascriptional promoter sequence whitin the β repeats using the Neural Network Promoter Prediction tool. The promoter sequence (5'TGTCAAACAGTGCTTAATAAGGATTTGACAGGGTTTTTTAGGGTAAGTAA -3') has a credibility score of 0.98 on 1.00 and is located 1 bp downstream the palindromic motif. As a matter of fact, the occurrence of a putative promoter sequence in each of the β repeats seems to suggest that these sequences may be spacer promoters or trascriptional enhancers. This hypothesis is also supported by the closeness of the palindromic motif to the putative promoter.

In fact, the local fold symmetry created by the palindrome is thought to provide the binding site for DNA-binding proteins involved in the trascriptional machinery (Arunkumar and Nagaraju, 2006).

In more detail, palindromic motifs (and in general inverted repeats) have the potential to form single-stranded stem-loop cruciform structures which have been reported to be essential for replication of circular genomes in many prokaryotic and eukaryotic systems (Cheung, 2004).

Nonetheless, the presence of spacer promoters/trascriptional enhancers in combination with the mechanism leading to the omogenization of the tandemly repeats, above all in male-LUR, strongly resemble the structure of IGS (intergenic spacer) in rDNA. In fact, the IGS contains the RNA-PolI promoter and other important regulatory elements, such as terminators, spacer promoters and enhancers to form clusters of regulatory subrepeats and the redundancy of those functional elements (repeats) (Moss, 1995; Reeder, 1999 has been related to an increased rate of rDNA.

Given that what we above reported is true, a higher number of regulatory elements are present in *Musculista* male LUR, when compared to that of the female. We therefore can speculate that this may point out an explanation of how sperm mitochondrial DNA becomes the dominant or the exclusive mtDNA component of the male gonads. Actually, it is known from early electron microscopy studies (Longo and Dornfeld 1967) that the midpiece of *Mytilus* sperm at maturity normally carries five mitochondria that are much larger than egg mitochondria. As in most other animals, the sperm mitochondria penetrate the cell membrane of the ovum at fertilization. However, paternal mtDNA is a minority in the zygote at fertilization, as the egg contains many tens of

thousands of F-type mitochondria (Humphreys, 1962). Thus, at fertilization, all embryos have a very large bias in favor of the F genome. Clearly, in species with DUI, different mechanisms of mitochondrial replication and/or destruction must operate to produce adult females in which the male genome has largely disappeared and adult males in which somatic cells are primarily or entirely F-type but the gonads are M-type.

Actually, in DUI species, evidences have been found that sperm-derived mitochondria are actively segregated to germ cells in male embryos: in the DUI system of *Mytilus*, the fate of sperm mitochondria is different if the embryo will develop a male or a female, since in the first case, sperm mitochondria tend to aggregate in a single blastomere, likely the precursor of the male germline, while in female embryos they are dispersed and quickly degraded (Cao et al., 2004; Obata and Komaru, 2005; Cogswell et al., 2006).

On the other hand, the question of how sperm mitochondrial DNA becomes dominant or the exclusive component of the male gonad remains open. In principle, two alternatives have been proposed: the first is that only sperm mitochondria gain entrance into the first germ cells; the second is that both sperm and egg mitochondria enter the primordial cells, but sperm mitochondrial DNA enjoys a replication advantage over the egg mitochondria DNA in these cells. As a general remark, male-LUR of *Musculista senhousia* would seem point out this latter interpretation as the most credible: a higher number of regulatory elements would advantage male mtDNAs over the female ones, because of the resulting increase of replication rate in the early developmental stages.

## *3.2 PHASMATODEA*

### *3.2.1 RESULTS*

The partial mtDNA genomes, including the region from *tRNA-Met* to *rrnS* genes of *Bacillus atticus* and *Bacillus rossius* stick insect (order Phasmatodea, suborder Verophasmatodea), were sequenced for this study. The sequenced region include all the protein coding genes and are 14,152 bp and 14,138 bp long in *B. atticus* and *B. rossius* respectively. I was unable, as it was for *Timema californicum* stick insect (Cameron et al., 2006), to successfully sequence the control region in *Bacillus*. As mentioned, such a failure should be either due to its extreme length or to the presence of highly repetitive A+T-rich portions in this region, or to both.

The mtDNA genome of both *B. atticus* and *B. rossius* possessed the typical metazoan mitochondrial genome composition of 13 protein-coding genes, 2 ribosomal RNAs and 19 out of 22 transfer RNAs. Moreover, the observed gene orders are identical to that proposed by Boore (1999) as ancestral arrangement (synplesiomorphic) for insects.

The overall AT-content are 78.1% and 77.6% in *B. atticus* and *B. rossius,* respectively. As in typical arthropod mtDNA, there are only small non-coding regions between genes: smaller non-coding regions are present between *nad5* and *trnH* (60 bp), between *trnS2* and *nad1* (65 bp) and between *trnP* and *nad6* (1 bp) (Tab.8) in *Bacillus rossius,* while in *Bacillus atticus* they are found between *nad5* and *trnH* (60 bp), *trnP* and *nad6* (1 bp), *trnS2* and *nad1* (67 bp), *trnR* and *trnN* (1 bp), *trnN* and *trnS1* (1 bp), *trnH* and *nad4* (2 bp), *nad4L* and *trnT* (8 bp) and between *nad1* and *trnL1* (3 bp) (Tab.9).

As shown in Table9, the protein-coding genes in *Bacillus atticus* showed two different start codons: ATG (used 6 times) and ATA (used 7 times). Also *Bacillus rossius* (Tab.8) showed two different start codons, ATG (used 5 times) and ATA (used 8 times). In the *nad2*, *cox1*, *cox2*, *cox3*, *nad3* and *cob* genes, stop codons are represented by a T (truncated TAA), whereas the all the remaining are TAA (6 times), but one TAG (*nad1*) in both *B. atticus* and *B. rossius*.

As in all other mitochondrial genomes sequenced so far, two genes for ribosomal RNAs were present, *rrnL* and *rrnS*, one for the large and one for the small ribosomal subunit. All 19 out 22 tRNAs can be folded into typical cloverleaf secondary structures.

### *3.2.2 PHYLOGENETIC ANALYSES AND EFFECT OF DATA MANIPULATION*

Both ALL (MP-ALL, ML-ALL and BA-ALL), PCG12 (MP-PCG12, ML-PCG12 and BA-PCG12) and PROT (MP-PROT and BA-PROT) analyses constantly supported Phasmatodea (*Timema+Bacillus*) as monophyletic with high bootstrap or posterior probabilities values.

In contrast, different optimality criteria did produce different relationships between Phasmatodea and other ingroups: actually, phylogenetic analyses under heuristic parsimony indicated a variety of sister groups for Phasmatodea depending on the coding scheme used: Phasmatodea+Odonata (MP-ALL, with gaps coded as 5th state and 100% of bootstrap support); a politomy with Orthoptera, Dictyoptera (Mantodea+(Blattodea+Isoptera)), Mantophasmatodea and Grylloblattodea (MP-ALL, with gaps coded as missing data); a complete politomy with all the others ingroups (MP-PCG12); and Phasmatodea+Mantophasmatodea+Grylloblattodea+Dictyoptera (MP-PROT), with high values of bootstrap.

In contrast, Bayesian and Maximum Likelihood analyses were more consistent for both tree topology and nodal support, regardless of the optimality criteria used. Constantly, a sister-group relationship was found (Phasmatodea+(Mantophasmatodea+Grylloblattodea)) with this clade, always closely related to Dictyoptera.

Nodal support values also appeared to be affected by the optimality criteria. In other words, the bootstrap values in MP analyses were generally lower than the bootstrap values in ML and posterior

probabilities in BA analyses. However, there was an exception to this general pattern, found in the MP-ALL (with gaps coded as 5th state) where node subtending Phasmatodea had a bootstrap value of 99% that is in line with and BA posterior probability (1.00) and considerably higher than ML bootstrap values.

Inclusion and exclusion of data had a greater influence in analyzing mitochondrial genome both in topology and nodal support. I compared the effect of partitioning according to gene type (PCG, TRAN, RIBO) against the total evidence dataset (ALL) and protein-coding genes only (PCG).

When transfer RNAs were analyzed as a single partition (TRNA), a monophyletic Phasmatodea was recovered in two of the four analyses (MP-TRNA with gaps coded as missing data and BA-TRNA although with 52% of bootstrap value and posterior probabilities of 0.96 respectively).

Also the analyses based on ribosomal RNAs (RIBO) poorly resolved the relationships among clades if compared to the combined dataset, resulting in unique and incongruent topologies from each of the different inference methods. Only the MP-RIBO (with gaps coded as 5th state) analysis produced a monophyletic Phasmatodea (bootstrap value 94%).

Both RIBO and TRNA coding schemes failed to recover any type of relationship with the others ingroups. Even the analyses based on combined structural and ribosomal RNAs (RIBO+TRNA) failed to resolve relationships compared to the combined dataset: only in the case of heuristic parsimony (MP-RIBO+TRNA) with gaps coded as 5th state, a monophyletic Phasmatodea clade is recovered with 94% of nodal support. Moreover, under this coding scheme, I found Phasmatodea clade to be the sister group of Odonata, forming a clade strongly supported (100% of bootstrap value).

Different methods of gap coding had varying degrees of influence in phylogenetic reconstruction, especially in relation to the size of dataset. In fact, only ALL scheme with gaps coded as missing data, supported the basal Ephemeroptera hypothesis ((Ephemeroptera versus (Odonata+Neoptera)), while all the other schemes did not supported this. regardless the method of gaps coding.

In all elaborations done on the large datasets (ALL and PCG), Phasmatodea remained a well defined monophyletic group, as well as Dictyoptera, regardless the method of gaps coding. Interestingly in ALL scheme with gaps coded as 5th state, Odonata resulted to be the unique sister group to Phasmatodea. Also in all small datasets (TRNA, RIBO and TRNA+RIBO), I observed a considerable effect of gap coding. On the whole, none of the schemes supported the basal Ephemeroptera hypothesis. The Mantophasmatodea clade results to be the sister group to (Paleoptera+Neoptera) in both RIBO and RIBO+TRNA schemes when gaps were coded as 5th state. When gaps were coded as missing data, monophyletic Phasmatodea were recovered only in TRNA scheme and no sister-relationship with any other polyneopteran order is recovered.

Phasmatodea resulted monophyletic also in RIBO and RIBO+TRAN schemes when gaps were coded as 5th state. As already observed, in ALL scheme with gaps coded as 5th state, Odonata resulted to be unique sister group to Phasmatodea.
The monophyly of Phasmatodea was found also when amino acid sequences were analyzed (PROT), regardless the method of gaps coding. PROT schemes strongly supported the relationship (Phasmatodea+Mantophasmatodea+Grylloblattodea) with this clade that is sister to Dictyoptera.
On the whole, even PROT schemes rejected the basal Ephemeroptera hypothesis with Orthoptera moving to be sister to Odonata.

### *3.2.3 DISCUSSION*

Both the newly sequenced mitochondrial genomes are similar in gene and nucleotide composition to that of *Timema californicum* stick insect mitochondrial genome (Cameron et al., 2006), as well as to the presumed ancestral hexapod (Boore, 1999; Fenn et al., 2007; Kim et al., 2005).
In the last decade, mitochondrial gene rearrangements was considered potentially useful phylogenetic markers and have been shown to be useful in resolving deep level relationships within Arthropoda (Boore et al., 1998). In contrast, as additional mt genomes have been sequenced, the phylogenetic utility of gene rearrangement in understanding insect evolution now appears to be minor, because most insect mt genomes have retained the ancestral gene order and it would not seem to be enough to resolve inter-Order or deeper relationships. Nonetheless, exceptions do exist, with rearrangements being found in insect orders such as Hymenoptera (Castro et al., 2006; Crozier and Crozier, 1993), Thysanoptera (Shao and Barker, 2003), Phthiraptera (Cameron and Whiting, 2007; Covacin et al., 2006; Shao et al., 2001), and Orthoptera (Flook et al., 1995b).
This study expands previous knowledge of mitochondrial genome in Phasmatodea adding to the sampling (*Timema californicum*, suborder Timematodea) two newly representatives of the unstudied (to date) suborder Verophasmotodea.
Although this taxa sampling is too small to make definitive conclusion about all aspects of Phasmatodea evolution, it is interesting to note that the ancestral plesiomorphic gene order is shared by both Timematodea and Verophasmatodea suborders, with the former that is the earliest branching group of the Phasmatodea order.
The possibility that these gene order may represent secondary reversion to the insect ground plan is therefore low and nonetheless, it is doubtful that additional intraordinal synapomorphies may come to the light.
The total evidence datasets (ALL coding scheme), regardless of the optimality criteria used, all

recovered Phasmatodea as monophyletic. This is particularly interesting because the Timematodea suborder is the earliest diverging stick insect. Actually it has been shown that this suborder is the sister to the Euphasmatodea (to which Verophasmatodea belong) in previous phylogenetic studies (Whiting et al., 2003). Moreover the divergence between *Timema* and Euphasmatodea occurred more than 95 Myr ago (Buckley et al., 2009 in press).

Under Maximum Likelihood and Bayesian analyses, a quite robust sister relationship among (Phasmatodea+(Mantophasmatodea+Grylloblattodea)) is found, with this clade more closely related to Dictyoptera ((Mantodea+(Blattodea+Isoptera) rather than Orthoptera. Thus the inferred trees allow to reject the "robust" sister relationship found by Fenn et al., (2008), which was rather concordant with many previous studies that found a close relationship between Orthoptera and Phasmatodea (Sharov, 1968; Kamp, 1973; Boudreaux, 1979; Kukalová-Peck, 1991; Flook and Rowell, 1998; Wheeler et al., 2001; Terry and Whiting, 2005; Cameron et al., 2006).

Complete taxon sampling of rich character sets will eventually lead to a more accurate understanding of the phylogenetic relationships within Polyneoptera and the sequencing of complete mtgenomes from additional representatives will aid in this effort.

Another interesting issue can be mentioned for the clade ((Phasmatodea+(Mantophasmatodea/Grylloblattodea). Actually, the total evidence data (ALL) alone are not enough to unambiguously resolve whether Mantophasmatodea is more closely related to Phasmatodea rather than to Grylloblattodea, as supposed by Terry and Whiting (2005). In more detail, the morphological synapomorphies defining the polyneopterous orders need to be rigorously assessed to reconcile molecular studies with traditional morphological assessments: Phasmatodea is defined by several synapomorphies, such as pear-shaped secretory appendices on the posterior part of the mesenteron (Bradler, 2003), the absence of mitochondria in spermatozoa (Jamieson, 1987), the male vomer (Bradler, 1999, 2003), the splitting of the lateral dorsoventral musculature into isolated muscle fibres (Bradler, 2003) and the emarginated labrum (Tilgner et al., 1999). More attention has centred on the presence of paired prothoracic repellant glands (Bradler, 2003; Hennig, 1969, 1994; Tilgner et al., 1999), the absence of which in Mantophasmatodea is probably the basis of the failure to determine a close relationship between these two groups, in spite of the gross morphological similarities between mantophasmids and phasmatodeans in general and *Timema* in particular. Cameron et al., (2006) suggested two possibilities, either that this gland was secondarily lost in Mantophasmatodea, or that it is not a synapomorphy of an expanded Phasmatodea clade (including Mantophasmatodea) but rather, a synapomorphy of a more restricted set of taxa. The analysis of genitalic characters (Klass et al., 2003) points to affinity between Phasmatodea and Mantophasmatodea; on the other hand, further characters need to be examined more in detail to

support this conclusion.
As a general remark, deeper analyses of the relationships between *Timema* and other early branching stick insects need to be performed in order to draw more rigorous conclusions.

#### *3.2.3.1 METHODOLOGICAL EFFECTS OF VARIOUS APPROACHES TO PHYLOGENETIC RECONSTRUCTION*

The differences, advantages and disadvantages between the three major phylogenetic optimality criteria (MP, ML and BA) are an ongoing debate. In this study, I assessed how each criterion would perform under a variety of different datasets derived from the mt genomes. When all available data were analyzed simultaneously (ALL-12), there was no apparent effect on topology between parsimony and model-based analyses. Even when a unique model of molecular evolution was applied to a dataset composed of multiple genes with differing evolutionary rates (ML), topology was the same as the highly precise and partitioned evolutionary models (BA) and high nodal supports were still recovered. These findings may indicate that the signals within mt genomes are so strong that even inadequate models of evolution do not adversely affect the phylogenetic reconstruction. However, it is also possible that the strong congruence among different optimality criteria is a function of small taxon sampling and large character sampling. In fact, when the smaller subsets of data (RIBO or TRAN) are analyzed under different optimality criteria, different topologies were recovered. In fact, the great part of the analyses using ribosomal RNAs and transfer RNAs failed to recover Pasmathodea as monophyletic. In contrast, they do contribute to the overall signal when analyzed with other available data. Especially, we find that the TRAN dataset had a comparable resolving capacity to the RIBO dataset, despite its smaller size of being approximately 50% of the RIBO dataset.

# 4. MATERIALS AND METHODS

## *4.1 SECTION-1: MOLLUSCA BIVALVIA*

### *4.1.1 SAMPLE COLLECTION*

*Musculista senhousia* specimens from Venice Lagoon (Italy) were used for analysis. About 50 specimens were stimulated to emit sperm or eggs in seawater added with hydrogen peroxide, according to (Morse et al., 1997). As soon as the treated seawater was removed and the mussels introduced in single recipients with clear seawater, some of them started emitting sperm or eggs. Each emission was analyzed by light microscopy to sex the specimens, as well as to detect eventual contamination by somatic cells, and a total of 10 sperm and 10 egg samples were further analyzed. Gametes were then collected after a gentle centrifugation (3000 × g) and seawater removed. Gametes were immediately frozen and stored at -20°C for subsequent analyses.

### *4.1.2 PCR AMPLIFICATIONS, SEQUENCING AND ANNOTATION*

Total genomic DNA was extracted from each gamete sample using the DNeasy Tissue Kit (Qiagen).

Partial sequences of Cytochrome b (*cob*) and mitochondrial ribosomal large subunit RNA (*rrnL*) were amplified and directly sequenced without cloning, as described in (Passamonti et al., 2003) The primers were: cobR (5'-GCRTAWGCRAAWARRAARTAYCAYTCWGG-3') and cobF (5'-GGWTAYGTWYTWCCWTGRGGWCARAT-3') for *cob* (designed by J. L. Boore), and 16Sbr and 16SarL for *rrnL* (Palumbi et al., 1991). Sequencing reactions were performed on both strands with BigDye Terminator Cycle Sequencing Kit according to supplier's instructions (Applied Biosystem) in a 310 Genetic Analyzer (ABI) automatic sequencer.

The sequences obtained from these region were then used to design sex-specific primers to amplify the entire mitochondrial genome in two overlapping fragments by long PCR reactions.

Long PCR amplifications were on a Gene Amp® PCR System 2720 (Applied Biosystem) in 50 $\mu$l reaction volume composed of: 31.5 $\mu$l of sterilized distilled water, 10 $\mu$l of 5X Herculase II Fusion Reaction Buffer, 0.5 $\mu$l of dNTPs mix, 1.25 $\mu$l of each primer (10 $\mu$M), 5 $\mu$l of DNA template (25-50 ng) and 0.5 $\mu$l of Herculase II Fusion DNA Polymerase.

Reaction conditions were in accordance to supplier's recommendations: initial denaturation at 95°C for 5 min and then incubated at 95°C for 20 sec, 50°C for 20 sec, and 68°C for 10 min for 30 cycles and 68°C for 8 min for a final extension. To obtain the F genome, the mtDNA of a female individual was amplified by long PCR in two fragments: both pairs of primers, F-cob383R//F-16S142F and F-cob160F//F-16S355R, amplified a fragment of 10-11 kb respectively.

The M genome of one male individual was amplified by long PCR in two fragments using the same reaction conditions and two sets of primers: M-cob386R//M-16S103F and M-cob160mF//M-16S218R. Both couple of primers amplified a fragment of 10-11 kb respectively.

Each Long-PCR fragments were purified using Wizard® SV Gel and PCR Clean-Up System (Promega).

Sequencing of the two major fragments was done using a shotgun approach. Amplicons were randomly sheared to 1.2–1.5 kb DNA segments using a HydroShear device (GeneMachines). Sheared DNA was blunt end-repaired at room temperature for 60 min using 6 U of T4 DNA Polymerase (Roche), 30 U of DNA Polymerase I Klenow (NEB), 10 $\mu$l of dNTPs mix, 13 $\mu$l of 10× NEB buffer 2 (NEB) in a 115 $\mu$l total volume and gel purified using the Wizard® SV Gel and PCR Clean-Up System (Promega). The resulting fragments were ligated into the SmaI site of a pUC18 cloning vector using the Fast-Link DNA ligation Kit (Epicentre) and electroporated into One Shot® TOP10 Electrocomp™ *E. coli* cells (Invitrogen) using standard protocols. Recombinant clones were screened by PCR using M13 universal primers. Obtained recombinant colonies were purified using Multiscreen (Millipore) according to the manufacturer's instructions. Clones were sequenced using M13 universal primers by Macrogen Inc.

### *4.1.3 ANNOTATION, ALIGNMENT AND PHYLOGENETIC ANALYSES*

Raw sequences were manually corrected and assembled into contigs with the software Sequencher 4.6 (Gene Codes). Hence, the final assemblies were based on a minimum sequence coverage of 3X.

The tRNA genes were identified by their secondary structure using tRNA-scan SE 1.21 (Lowe and Eddy, 1997) with invertebrate mitochondrial codon predictors and a cove score cut off of 1. Open reading frames between tRNAs were found using ORF Finder and identified using translated BLAST searches (blastx) (Altschul et al., 1997) as both implemented by the NCBI website (http://www.ncbi.nlm.nih.gov/).

Amino acid alignment was generated in ClustalW (Thompson et al., 1994 implemented in MEGA4.0, Tamura et al. 2007) for each of the 13 protein coding genes and a DNA alignment

inferred from the amino acid alignment using MEGA4.0 (Tamura et al. 2007) which can translate between DNA and amino acid sequences within alignments.

To analyze sequence variability, pairwise p-distances, their mean values and standard errors (by the bootstrap procedure) were obtained using the software MEGA version 4.1 (Tamura et al. 2007). The use of a p-distance estimator has been preferred for simplicity, because I did not want to introduce any model of DNA substitution, which might have an influence in the performed tests, and because the use of p-distance estimators has been already used also in earlier literature (Passamonti et al., 2003).

The divergence of protein genes in synonymous (Ks) and nonsynonymous (Ka) sites was calculated by the modified Nei-Gojobori method with Jukes-Cantor correction, and the p distance at the amino acid level was calculated using the computer program MEGA version 4.1 (Tamura et al. 2007).

Twofold-, and fourfold-degenerated positions were identified using software DnaSP program v.4.1 (Rozas et al., 2003). This program was also used to estimate codon usage.

Potential secondary structures near or at the 5'-end of protein genes have been produced by the RNA "mfold" program 3.1 (Zuker 2003).

The Neural Network Promoter Prediction tool (Reese, 2001) was used to identify potential promoter sequences.

## *4.2 SECTION-2: PHASMATODEA*

### *4.2.1 SAMPLE COLLECTION*

Stick insects of *Bacillus rossius* and *Bacillus atticus* were collected from Sardinia island (Siniscola) and Israel (Golan) respectively. Field-collected specimens were stored at -80°C.

### *4.2.2 PCR AMPLIFICATIONS AND SEQUENCING*

Total genomic DNA was isolated from somatic tissues with a standard phenol-chloroform protocol. In each of the two specimens four partially overlapping pieces were obtained using universal primers: the fragments of *rrnS* gene (543 bp) were amplified using the pair of primers SR-J14197 and SR-N14745 (Simon et al. 2006) via normal PCR and directly sequenced. Partial genomes including the region *nadh2* to *cox1* genes (2100 bp) were amplified with primers TM-J210 (Simon et al. 1994) and C1-N2329 (Simon et al. 2006) via Long PCR and directly sequenced using “primer

walking" method. The two major fragments (9.0 kb and 5.5 kb) were amplified using C1-J-2195//CB-N-11367 and N4-J-8944//LR-N primers (Simon et al. 1994) respectively.

Normal PCR were performed in a 50 µL reaction mixture consisting of 27.5 µl of sterilized water; 3 µl MgCl2, 50 mM; 5 µl 10 × PCR Buffer; 4 µl dNTP, 2.5 mM; 2.5 µl of each primer, (10 µM); 5 µl DNA template (25-50 ng); 0.5 µl Takara Taq DNA polymerase: initial denaturation for 2 min at 94 °C, followed by 30 cycles of 30 s at 94 °C, 30 s at 52 °C, and 60 s at 72°C, and a subsequent 7 min final extension step at 72 °C.

Long PCR amplifications were in 50 $\mu$l reaction volume composed of: 31.5 $\mu$l of sterilized water, 10 $\mu$l of 5X Herculase II Fusion Reaction Buffer, 0.5 $\mu$l of dNTPs mix, 1.25 $\mu$l of each primer (10 $\mu$M), 5 $\mu$l of DNA template (25-50 ng) and 0.5 $\mu$l of Herculase II Fusion DNA Polymerase. Reaction conditions were according to supplier's recommendations: the mix was heated at 95°C for 5 min and then incubated at 95°C for 20 sec, 50°C for 20 sec, and 68°C for 10 min for 30 cycles and 68°C for 8 min for a final extension.

Both Normal- and Long-PCR were performed using Gene Amp® PCR System 2720 (Applied Biosystem)

As well as for *Timema californicum* stick insect (Cameron et al., 2006), it was not possible to successfully sequence the control region of both *Bacillus* specimens. This failure should be due to either its extreme length or to the presence of highly repetitive A+T-rich portions in this region, or both.

Each PCR fragments were purified using Wizard® SV Gel and PCR Clean-Up System (Promega).

Sequencing of the two major fragments was done using a shotgun approach. Amplicons were randomly sheared to 1.2–1.5 kb DNA segments using a HydroShear device (GeneMachines). Sheared DNA was blunt end-repaired at room temperature for 60 min using 6 U of T4 DNA Polymerase (Roche), 30 U of DNA Polymerase I Klenow (NEB), 10 $\mu$l of dNTPs mix, 13 $\mu$l of 10× NEB buffer 2 (NEB) in a 115 $\mu$l total volume and gel purified using the Wizard® SV Gel and PCR Clean-Up System (Promega). The resulting fragments were ligated into the SmaI site of a pUC18 cloning vector using the Fast-Link DNA ligation Kit (Epicentre) and electroporated into One Shot® TOP10 Electrocomp™ *E. coli* cells (Invitrogen) using standard protocols. Recombinant clones were screened by PCR using M13 universal primers. Obtained recombinant colonies were purified using Multiscreen (Millipore) according to the manufacturer's instructions. Clones were sequenced using M13 universal primers by Macrogen Inc.

### *4.2.3 ANNOTATION, ALIGNMENT AND PHYLOGENETIC ANALYSES*

Raw sequences were manually corrected and assembled into contigs with the software Sequencher 4.6 (Gene Codes). Hence, the final assemblies were based on a minimum sequence coverage of 3X. The tRNA genes were identified by their secondary structure using tRNA-scan SE 1.21 (Lowe and Eddy, 1997) with invertebrate mitochondrial codon predictors and a cove score cut off of 1. Open reading frames between tRNAs were found using ORF Finder and identified using translated BLAST searches (blastx) (Altschul et al., 1997) as both implemented by the NCBI website (http://www.ncbi.nlm.nih.gov/).

To elucidate the phylogenetic position of Verophasmatodea within pterygote insects, mtDNA sequences for 12 additional species of insects were obtained from GenBank two nonpterygotes, a silverfish (*Nesomachilis australica*) and a bristletail (*Tricholepidion gertischi*), were used as outgroup taxa (Tab. 10).

Annotated mt genomes were organized using MEGA 4.0 (Tamura et al. 2007) with each of the 13 protein coding genes aligned separately. Protein-coding genes were translated into amino acid sequence using the invertebrate mitochondrial genetic code in MEGA, and aligned based on their amino acid sequence using default settings in ClustalW (as implemented in MEGA). The alignment was back-translated into the corresponding nucleotide sequences. Ribosomal and transfer RNA genes (19 out 22: except for *trnI*, *trnQ* and *trnM*) were aligned individually with MAFFT. Datasets were concatenated in MacClade (Maddison and Maddison, 2003) using the partitioning scheme presented below.

### *4.2.4 PHYLOGENETIC ANALYSES AND DATA PARTITIONING*

An alignment consisting of 3.835 amino acids and 11.505 nucleotides were obtained from all protein genes. A larger alignment of 15.238 nucleotides was obtained by adding to the 11.505 nucleotides both transfer and ribosomal RNA genes (1.253 and 2.480 nucleotides respectively).

For each coding protein gene, a saturation analysis (Xia et al., 2003) was performed on third codon position using DAMBE 4.2.13 (Xia and Xie, 2001). Most of the genes showed saturation on third codon positions, thus they were excluded from the final concatenated nucleotide alignment and an alignment of 7670 nucleotides for protein coding genes was obtained.

The effect of different inference methods on topology and nodal support in mt genome phylogenies has been tested using parsimony (MP), maximum likelihood (ML) and Bayesian (BA) analyses (Figs. 7-9 for all trees obtained). MP analyses were performed using PAUP* ver. 4.0b10 (Swofford,

2002), with gaps treated both as missing or as a 5th character state (21st character state for amino acid analyses). Bootstrap support was calculated from 1000 bootstrap replicates with 100 random additions per replicate in PAUP*. Tree statistics were also calculated in PAUP* (Table x). Models for ML analyses were chosen for each concatenated dataset using "Hierarchical Likelihood Ratio Tests" (hLRTs) as implemented in ModelTest (Posada and Crandall, 1998). Bootstrap support for ML trees was calculated using 500 bootstrap replicates with 10 random additions per replicate. Models for BA analyses were also chosen using AIC as implemented in ModelTest (Posada and Crandall, 1998), with models assessed independently for each partition used in the concatenated datasets (see below). All BA analyses ran four separate runs using unlinked partitions with four chains per run for a total of ten million generations per run with sampling every 100 generations in MrBayes versions 3.1.1 (Ronquist and Huelsenbeck, 2003).

Parameter values were examined for asymptotic behavior and all generations prior to this point were discarded as burn-in: therefore, an average of 10% of each BA run was treated as burn-in and discarded. All phylogenetic analyses were run on the BIOportal server (www.bioportal.uio.no).

I also studied the effect of data partitioning by comparing topology and nodal support values within each of three phylogenetic inference methods tested. In this regard, six datasets have been created with varying gene content or coding: *i)* all genes, excluding the third codon position of the protein-coding genes (protein-coding, ribosomal RNA and transfer RNA genes), coded as nucleotide data (ALL); *ii)* protein-coding genes excluding third codon positions alone coded as nucleotides (PCG-12) ; *iii)* protein-coding genes translated into amino acid sequences (PROT); *iv)* transfer RNA genes alone (TRNA); *v)* ribosomal RNA genes alone (RIBO) and, *vi)* transfer RNA together with ribosomal genes (TRNA+RIBO). In this way, it was possible to study the phylogenetic information content of different gene types (ALL vs PCG-12 vs RIBO vs TRNA vs TRNA+RIBO datasets) and the effects of translated amino acids versus nucleotide sequences (ALL/PCG-12 vs PROT datasets). Additionally, the effect of gap coding was explored in MP analyses by coding gaps as missing vs as a new character state (5th state in nucleotide sequences, 21st state in amino acid sequence). In BA analyses, a single partition has been used for TRNA dataset. Differently, the ALL, PCG-12, RIBO and TRNA+RIBO datasets have been partitioned into gene-based partitions (GP, 16 partitions in ALL, 13 in PCG-12, 2 in RIBO and 3 in TRNA+RIBO, to examine the effect of different partitioning schemes used in BA analyses.

## 5. REFERENCES


Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389–3402.

Arunkumar KP, Nagaraju J (2006). Unusually long palindromes are abundant in mitochondrial control regions of insects and nematodes. PLoS ONE 1: e110.

Asakawa S, Kumazawa Y, Araki T, Himeno H, Miura K, Watanabe K (1991). Strand-specific nucleotide composition bias in echinoderm and vertebrate mitochondrial genomes. J Mol Evol 32: 511–520.

Bae JS, Kim HD, Sohn BR (2004). The mitochondrial genome of the firefly *Pyrocoelia rufa*: complete genome sequence, genome organisation and phylogenetic analysis with other insects. Mol Phyl Evol 32: 978-985.

Ballard JWO (2000a). Comparative genomics of mitochondrial DNA in *Drosophila simulans*. Journal of Molecular Evolution 51: 48–63.

Beagley CT, Okimoto R, Wolstenholme DR (1999). *Mytilus* mitochondrial DNA contains a functional gene for a tRNASer(UCN) with a dihydrouridine arm-replacement loop and a pseudo-tRNASer(UCN) gene. Genetics 152: 641–652.

Bibb MJ, Van Etten RA, Wright CT, Walberg MW, Clayton DA (1981). Sequence and gene organization of mouse mitochondrial DNA. Cell 26: 167–180.

Bielawski JP, Gold JR (2002). Mutation patterns of mitochondrial H- and L-strand DNA in closely related Cyprinid fishes. Genetics 161: 1589–1597.

Boore JL (1999). Animal mitochondrial genomes. Nucleic Acids Res 27: 1767-80.

Boore JL (2000). The duplication/random loss model for gene rearrangement exemplified by

mitochondrial genomes of deuterostome animals. In: Sankoff D, Nadeau JH (eds). Comparative Genomics, 1st edn, vol. 1. Kluwer Academic: Dordrech/Boston/London. pp 133–147.

Boore JL, Brown WM (1994). Complete DNA sequence of the mitochondrial genome of the black chiton *Katharina tunicata*. Genetics 138: 423–443.

Boore JL, Brown WM (1994). Mitochondrial genomes and the phylogeny of molluscs. *Nautilus* 108 (suppl. 2); 61-78.

Boore JL, Brown WM (1995). Complete sequence of the mitochondrial DNA of the annelid worm *Lumbricus terrestris*. Genetics 141: 305–319.

Boore JL, Brown WM (1998). Big trees from little genomes: mitochondrial gene order as a phylogenetic tool. Curr Opin in Genet & Develop 8: 668–674.

Boore JL, Brown WM (2000). Mitochondrial genomes of *Galathealinum*, *Helobdella*, and *Platynereis*: sequence and gene arrangement comparisons indicate that Pogonophora is not a phylum and Annelida and Arthropoda are not sister taxa. Mol Biol Evol 17: 87–106.

Boore JL, Medina M, Rosenberg LA (2004). Complete sequences of the highly rearranged molluscan mitochondrial genomes of the scaphopod *Graptacme eborea* and the bivalve *Mytilus edulis*. Mol Biol Evol 21: 1492–1503.

Boore JL, Lavrov DV, Brown WM (1998). Gene translocation links insects and crustaceans. Nature 392: 667–668.

Borras C, Sastre J, Garcia-Sala D, Lloret A, Pallardo FV, Vina J (2003). Mitochondria from females exhibit higher antioxidant gene expression and lower oxidative damage than males. Free Radic Biol Med 34: 546-52.

Boyce TM, Zwick ME, Aquadro CF (1989). Mitochondrial DNA in the bark weevils: Size, structure and heteroplasmy. Genetics 123: 825-836.

Boudreaux HB (1979). Arthropod Phylogeny. John Wiley & Sons, New York.

Bradler S (1999). The vomer of *Timema* Scudder, 1895 (Insecta: Phasmatodea) and its signifcance for phasmatodean phylogeny. Cour. Forschungstinst. Senckenb. 215, 43–47.

Bradler S (2003). Phasmatodea, Gespentschrecken. In: Dathe, H.H. (Ed.), Lehrbuch der speziellen Zoologie. I. Wirbellose Tiere. 5. Insecta. Spektrum. Heidelberg, Berlin, pp. 251–260.

Breton S, Beaupre HD, Stewart DT, Hoeh WR, Blier PU (2007). The unusual system of doubly uniparental inheritance of mtDNA: isn' t one enough? Trends Genet 23: 465–474.

Breton S, Burger G, Stewart DT, Blier PU (2006). Comparative analysis of gender-associated complete mitochondrial genomes in marine mussels (*Mytilus* spp.). Genetics. 172: 1107-19.

Bridge D, Cunningham CW, Schierwater B, Desalle R, Buss LW (1992). Class-level relationships in the phylum Cnidaria: evidence from mitochondrial structure. Proc Natl Acad Sci USA 89: 8750-8753.

Brookes PS, Levonen AL, Shiva S, Sarti P, Darley-Usmar VM (2002). Mitochondria: regulators of signal transduction by reactive oxygen and nitrogen species. Free Radic Biol Med 33: 755-64.

Broughton RE, Milam JE, Roe BA (2001). The complete sequence of the zebrafish (*Danio rerio*) mitochondrial genome and evolutionary patterns in vertebrate mitochondrial DNA. Genome Res 11: 1958–1967.

Broughton RE, Reneau PC (2006). Spatial covariation of mutation and nonsynonymous substitution rates in vertebrate mitochondrial genomes. Mol Biol Evol 23: 1516–1524.

Brugler MR, France SC (2007). The complete mitochondrial genome of the black coral *Chrysopathes formosa* (Cnidaria: Anthozoa:Antipatharia) supports classification of antipatharians within the subclass Hexacorallia. Mol Phyl Evol 42: 776–788.

Buckley TR, Attanayake D, Bradler S (2009). Extreme convergence in stick insect evolution: phylogenetic placement of the Lord Howe Island tree lobster. Proc R Soc B: in press.

Burger G, Forget L, Zhu Y, Gray MW, Lang BF (2003). Unique mitochondrial genome architecture

in unicellular relatives of animals. Proc Natl Acad Sci USA 100: 892–897.

Burzynski A, Zbawicka M, Skibinski DO, Wenne R (2003). Evidence for recombination of mtDNA in the marine mussel *Mytilus trossulus* from the Baltic. Mol Biol Evol 20: 388-92.

Burzynski A, Zbawicka M, Skibinski DO, Wenne R. 2006. Doubly uniparental inheritance is associated with high polymorphism for rearranged and recombinant control region haplotypes in Baltic *Mytilus trossulus*. Genetics 174: 1081–1094.

Cameron SL, Barker SC, Whiting MF (2006). Mitochondrial genomics and the new insect order Mantophasmatodea. Mol Phyl Evol 38: 274–279.

Cameron SL, Lambkin CL, Barker SC, Whiting MF (2007). A mitochondrial genome phylogeny of Diptera: whole genome sequence data accurately resolve relationships over broad timescales with high precision. Systematic Entomology 32: 40–59.

Cameron SL, Miller KB, D'Haese CA, Whiting MF, Barker SC (2004). Mitochondrial genome data alone are not enough to unambiguously resolve the relationships of Entognatha, Insecta and Crustacea sensu lato (Arthropoda). Cladistics 20: 534–557.

Cameron SL, Whiting MF (2007). Mitochondrial genomic comparisons of the subterranean termites from the Genus Reticulitermes (Insecta: Isoptera:Rhinotermitidae). Genome 50: 188–202.

Cantatore P, Roberti M, Rainaldi G, Gadaleta MN, Saccone C (1989). The complete nucleotide sequence, gene organization, and genetic code of the mitochondrial menome of *Paracentrotus lividus*. J Biol Chem 264: 10965–10975.

Cao L, Kenchington E, Zouros E. (2004) Differential segregation patterns of sperm mitochondria in embryos of the blue mussel (*Mytilus edulis*). Genetics 166: 883-94.

Carapelli A, Lio P, Nardi F, Van der Wath E, Frati F (2007). Phylogenetic analysis of mitochondrial protein coding genes confirms the reciprocal paraphyly of Hexapoda and Crustacea. BMC Evolutionary Biology 7: S8.

Castro LR, Dowton M (2005). The position of the Hymenoptera within the Holometabola as

inferred from the mitochondrial genome of *Perga condei* (Hymenoptera: Symphyta: Pergidae). Mol Phyl Evol 34: 469–479.

Castro LR, Dowton M (2007). Mitochondrial genomes in the Hymenoptera and their utility as phylogenetic markers. Systematic Entomology 32: 60–69.

Castro LR, Ruberu K, Dowton M (2006). Mitochondrial genomes of *Vanhornia eucnemidarum* (Apocrita: Vanhorniidae) and *Primeuchroeus* spp. (Aculeata: Chrysididae): evidence of rearranged mitochondrial genomes within the Apocrita (Insecta: Hymenoptera). Genome 49, 752–766.

Chakrabarti R, Walker JM, Stewart DT, Trdan RJ, Vijayaraghavan S, Curole JP et al (2006). Presence of a unique male-specific extension of C-terminus to the cytochrome c oxidase subunit II protein coded by the male-transmitted mitochondrial genome of *Venustaconcha ellipsiformis* (Bivalvia: Unionoidea). FEBS Lett 580: 862-6.

Chalwatzis N, Hauf J, Van de Peer Y, Kinzelbach R, Zimmermann RK (1996). 18S ribosomal-RNA genes of insects: primary structure of the genes and molecular phylogeny of the Holometabola. Ann Entomol Soc Am 89: 788–803.

Chan WY et al. (2006). Transcriptome analyses of male germ cells with serial analysis of gene expression (SAGE). Mol Cell Endocrinol 250: 8–19

Chang YS, Huang FL, Lo TB (1994). The complete nucleotide sequence and gene organization of carp (*Cyprinus carpio*) mitochondrial genome. J Mol Evol 38: 138–155.

Cheung AK (2004). Palindrome regeneration by template strand-switching mechanism at the origin of DNA replication of porcine circovirus via the rolling-circle melting-pot replication model. J Virol 78: 9016–9029.

Clary DO, Wolstenholme DR (1985). The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. J Mol Evol 22: 252–271.

Cogswell AT, Kenchington EL, Zouros E (2006). Segregation of sperm mitochondria in two- and four-cell embryos of the blue mussel *Mytilus edulis*: Implications for the mechanism of doubly

uniparental inheritance of mitochondrial DNA. Genome 49: 799-807.

Collins TM, Wimberger PH, Naylor GJP (1994). Compositional bias, character-state bias, and character-state reconstruction using parsimony. Syst Biol 43: 482–96.

Conklin EG (1897). The embryology of *Crepidula*. J Morphol 14: 1-226.

Covacin C, Shao R, Cameron S, Barker SC (2006). Extraordinary number of gene rearrangements in the mitochondrial genomes of lice (Phthiraptera: Insecta). Insect Molecular Biology 15: 63–68.

Crease TJ (1999). The complete sequence of the mitochondrial genome of *Daphnia pulex* (Cladocera: Crustacea). Gene 233: 89–99.

Crozier RH, Crozier YC (1993). The mitochondrial genome of the honeybee *Apis mellifera* complete sequence and genome organization. Genetics 133: 97–117.

Curole JP, Kocher TD (2002). Ancient sex-specific extension of the cytochrome c oxidase II gene in bivalves and the fidelity of doubly-uniparental inheritance. Mol Biol Evol 19: 1323-8.

Curole JP, Kocher TD (2005). Evolution of a unique mitotype-specific protein-coding extension of the cytochrome c oxidase II gene in freshwater mussels (Bivalvia: Unionoida). J Mol Evol 61: 381-9.

Dalziel AC, Stewart DT (2002). Tissue-specific expression of male-transmitted mitochondrial DNA and its implications for rates of molecular evolution in *Mytilus mussels* (Bivalvia: Mytilidae). Genome 45: 348-55.

De Giorgi C, Martiradonna A, Lanave C, Saccone C (1996). Complete sequence of the mitochondrial DNA in the sea urchin *Arbacia lixula*: conserved features of the echinoid mitochondrial genome. Mol Phyl Evol 5:323–332.

Dellaporta SL, Xu A, Sagasser S, Jakob W, Moreno MA, Buss LW et al. (2006). Mitochondrial genome of *Trichoplax adhaerens* supports placozoa as the basal lower metazoan phylum. Proc Natl Acad Sci USA 103: 8751–8756.

Dobzhansky T (1944). Chromosomal races in *Drosophila pseudoobscura* and *D. persimilis*. Pp. 47-145 in: T. Dobzhansky and C. Epling (eds.), Contributions to the Genetics, Taxonomy and Ecology of *Drosophila pseudoobscura* and its relatives. – Carnegie Inst. Of Wash: Publ 554

Dowton M, Castro LR, Austin AD (2002). Mitochondrial gene rearrangements as phylogenetic characters in the invertebrates: the examination of genome "morphology". Invertebrates Systematics 16: 345-356.

Dreyer H, Steiner G (2004). The complete sequence and gene organization of the mitochondrial genome of the gadilid scaphopod *Siphonodentalium lobatum* (Mollusca). Mol Phyl Evol 31: 605–617.

Dreyer H, Steiner G (2006). The complete sequences and gene organisation of the mitochondrial genomes of the heterodont bivalves *Acanthocardia tuberculata* and *Hiatella arctica* and the first record for a putative Atpase subunit 8 gene in marine bivalves. Front Zool 3: 13.

Elson JL, Andrews RM, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N (2001). Analysis of European mtDNAs for recombination. Am J Hum Genet 68: 145-53.

Elson JL, Lightowlers RN (2006). Mitochondrial DNA clonality in the dock: can surveillance swing the case? Trends Genet 22: 603–607.

Erpenbeck D, Voigt O, Adamski M, Adamska M, Hooper JN, Worheide G et al. (2007). Mitochondrial diversity of early-branching metazoa is revealed by the complete mt genome of a haplosclerid demosponge. Mol Biol Evol 24: 19–22.

Faith JJ, Pollock DD (2003). Likelihood analysis of asymmetrical mutation bias gradients in vertebrate mitochondrial genomes. Genetics 165: 735–745.

Faure E, Casanova JP (2006). Comparison of chaetognath mitochondrial genomes and phylogenetical implications. Mitochondrion 6: 258–262.

Fenn, JD, Cameron SL, Whiting MF (2007). The complete mitochondrial menome sequence of the Mormon cricket (*Anabrus simplex*: Tettigoniidae: Orthoptera) and an analysis of control region variability. Insect Molecular Biology 16: 239–252.

Fenn JD, Song H, Cameron SL, Whiting MF (2008). A preliminary mitochondrial genome phylogeny of Orthoptera (Insecta) and approaches to maximizing phylogenetic signal found within mitochondrial genome data. Mol Phyl Evol 49: 59–68.

Fernandez-Silva P, Enriquez JA, Montoya J (2003). Replication and transcription of mammalian mitochondrial DNA. Exp Physiol 88: 41-56.

Filipowicz M, Burzinsky A, Smietanka B, Wenne R (2008). Recombination in mitochondrial DNA of European mussels *Mytilus*. J Mol Evol 67: 377-388.

Fisher C, Skibinski DOF (1990). Sex-biased mitochondrial DNA heteroplasmy in the marine mussel *Mytilus*. Proc R Soc Lond B Biol Sci 242: 149-155.

Flook PK, Rowell CHF (1998). Inferences about orthopteroid phylogeny and molecular evolution from small subunit nuclear ribosomal DNA sequences. Insect Molecular Biology 7: 163–178.

Flook P, Rowell H, Gellissen G (1995a). Homoplastic rearrangements of insect mitochondrial transfer-RNA genes. Naturwissenschaften 82: 336–337.

Flook PK, Rowell CHF, Gellissen G (1995b). The sequence, organization, and evolution of the Locusta migratoria mitochondrial genome. J Mol Evol 41: 928–941.

Flot JF, Tillier S (2007). The mitochondrial genome of *Pocillopora* (Cnidaria: Scleractinia) contains two variable regions: the putative D-loop and a novel ORF of unknown function. Gene 401: 80–87.

Fuller KM, Zouros E (1993). Dispersed length polymorphism of mitochondrial DNA in the scallop *Placopecten magellanicus* (Gmelin). Curr Genet 23: 365-369.

Garesse R (1988). *Drosophila melanogaster* mitochondrial DNA: gene organization and evolutionary considerations. Genetics 118: 649–663.

Garrido-Ramos MS, Stewart DT, Sutherland BW, Zouros E (1998). The distribution of male-transmitted and female-transmitted mitochondrial DNA types in somatic tissues of blue mussels:

Implications for the operation of doubly uniparental inheritance of Mitochondrial DNA. Genome 41: 818–24.

Gaut BS, Lewis PO (1995). Success of maximum likelihood phylogeny inference in the four-taxon case. Mol Biol Evol 12: 152–62.

Giribet G, Wheeler WC (2002). On bivalve phylogeny: a high-level phylogeny of the mollusk class Bivalvia based on a combined analysis of morphology and DNA sequence data. Inv Biol 121: 271–324.

Gissi C, Iannelli F, Pesole G (2004). Complete mtDNA of *Ciona intestinalis* reveals extensive gene rearrangement and the presence of an *atp8* and an extra *trnM* gene in ascidians. J Mol Evol 58: 376–389.

Gissi C, Pesole G (2003). Transcript mapping and genome annotation of ascidian mtDNA using EST data. Genome Res 13: 2203–2212.

Gissi C, Iannelli F, Pesole G (2008). Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. Heredity 101: 301-20.

Gossling E (1992). Systematic and geographic distribution of *Mytilus.* In: Gossling E (ed). The Mussel *Mytilus*: Ecology, Physiology, Genetics and Aquaculture, 1st edn, vol. 1. Elsevier: Amsterdam. pp 1–20.

Gray MW, Lang BF, Cedergren R, Golding GB, Lemieux C, Sankoff D, Turmel M, Brossard N, Delage E, Littlejohn TG, Plante I, Rioux P, Saint-Louis D, Zhu Y, Burger G. (1998). Genome structure and gene content in protist mirochondrial DNAs. Nucleic Acids Res 15, 865-878.

Grande C, Templado J, Cervera JL, Zardoya R (2002). The complete mitochondrial genome of the nudibranch *Roboastra europaea* (Mollusca: Gastropoda) supports the monophyly of opisthobranchs. Mol Biol Evol 10: 1672–1685.

Grimaldi D, Engel MS. (2005). Evolution of the Insects. Cambridge, UK: Cambridge Univ. Press

Gruber KF, Voss RS, Jansa SA (2007). Base-compositional heterogeneity in the RAG1 locus among didelphid marsupials: implications for phylogenetic inference and the evolution of GC content. Syst Biol 56: 83–96.

Haen KM, Lang BF, Pomponi SA, Lavrov DV (2007). Glass sponges and bilaterian animals share derived mitochondrial genomic features: a common ancestry or parallel evolution? Mol Biol Evol 24: 1518–1527.

Hatzoglou E, Rodakis GC, Lecanidou R (1995). Complete sequence and gene organization of the mitochondrial genome of the land snail *Albinaria coerulea*. Genetics 140: 1353–1366.

Helfenbein KG, Brown WM, Boore JL (2001). The complete mitochondrial genome of the articulate brachiopod *Terebratalia transversa*. Mol Biol Evol 18: 1734–1744.

Helfenbein KG, Fourcade HM, Vanjani RG, Boore JL (2004). The mitochondrial genome of *Paraspadella gotoi* is highly reduced and reveals that chaetognaths are a sister group to protostomes. Proc Natl Acad Sci USA 101: 10639–10643.

Humphreys WJ (1962). Electron microscope studies on eggs of *Mytilus edulis.* J Ultrastruct Res 7: 467-487.

Hennig W (1969). Die Stammesgechichte der Insekten. Kramer, Frankfurt am Main.

Hennig W (1994). Wirbellose II. Gliedertiere. 5 AuX. (Nachdruck). Gustav Fischer, Jena.

Ho SYW, Jermiin LS (2004). Tracing the decay of the historical signal in biological sequence data. Syst Biol 53: 623–37.

Hoeh WR et al. (1991) Heteroplasmy suggests limited biparental inheritance of *Mytilus mitochondrial*-DNA. Science 251: 1488–1490

Hoeh WR, Stewart DT, Guttman SI (2002). High fidelity of mitochondrial genome transmission under the doubly uniparental mode of inheritance in freshwater mussels (Bivalvia: Unionoidea). Evolution Int J Org Evolution 56: 2252-2261.

Hoeh WR, Stewart DT, Saavedra C, Sutherland BW, Zouros E (1997): Phylogenetic evidence for role-reversals of gender-associated mitochondrial DNA in *Mytilus* (Bivalvia: Mytilidae). Mol Biol Evol 14: 959-967.

Hoeh WR, Stewart DT, Sutherland BW, Zouros E. (1996) Cytochrome c oxidase sequence comparisons suggest an unusually high rate of mitochondrial DNA evolution in *Mytilus* (Mollusca: Bivalvia). Mol Biol Evol. 13:418-21.

Hoffmann RJ, Boore JL, Brown WM (1992). A novel mitochondrial genome organization for the blue mussel, *Mytilus edulis*. Genetics 131: 397–412.

Iannelli F, Griggio F, Pesole G, Gissi C (2007a). The mitochondrial genome of *Phallusia mammillata* and *Phallusia fumigata* (Tunicata, Ascidiacea): high genome plasticity at intra-genus level. BMC Evol Biol 7: 155.

Ingman M, Kaessmann H, Paabo S, Gyllensten U (2000). Mitochondrial genome variation and the origin of modern humans. Nature 408: 708–713.

Jamieson BGM (1987). The Ultrastructure and Phylogeny of Insect Spermatozoa. Cambridge University Press, Cambridge.

Jeyaprakash A, Hoy MA (2007). The mitochondrial genome of the predatory mite *Metaseiulus occidentalis* (Arthropoda: Chelicerata: Acari: Phytoseiidae) is unexpectedly large and contains several novel features. Gene 391: 264–274.

Jermiin LS, Crozier RH (1994). The cytochrome b region in the mitochondrial DNA of the ant *Tetraponera rufoniger*: sequence divergence in Hymenoptera may be associated with nucleotide content. J Mol Evol 38: 282–294.

Jermiin LS, Foster PG, Graur D, Lowe RM, Crozier RH (1996). Unbiased estimation of symmetrical directional mutation pressure from protein-coding DNA. J Mol Evol 42: 476–480.

Jermiin LS, Graur D, Lowe RM, Crozier RH (1994). Analysis of directional mutation pressure and nucleotide content in mitochondrial cytochrome b genes. J Mol Evol 39: 160–173.

Jermiin LS, Ho SYW, Ababneh F, Robinson J, Larkum AWD (2004). The biasing effect of compositional heterogeneity on phylogenetic estimates may be underestimated. Syst Biol 53: 638–43.

Jiang ZJ, Castoe TA, Austin CC, Burbrink FT, Herron MD, McGuire JA et al. (2007). Comparative mitochondrial genomics of snakes: extraordinary substitution rate dynamics and functionality of the duplicate control region. BMC Evol Biol 7: 123.

Jin L, Nei M (1990). Limitations of the evolutionary parsimony method of phylogenetic analysis. Mol Biol Evol 7: 82–102.

Kamp JW (1973). Numerical classification of the orthopteroids, with special reference to the Grylloblattodea. Canadian Entomologist 105: 1235–1249.

Kenchington E, MacDonald B, Cao L, Tsagkarakis D, Zouros E (2002). Genetics of mother-dependent sex ratio in blue mussels (*Mytilus* spp.) and implications for doubly uniparental inheritance of mitochondrial DNA. Genetics 161: 1579-88.

Kim I, Cha SY, Yoon MH, Hwang JS, Lee SM, Sohn HD, Jin BR (2005). The complete nucleotide sequence and gene organization of the mitochondrial genome of the oriental mole cricket, *Gryllotalpa orientalis* (Orthoptera: Gryllotalpidae). Gene 353: 155–168.

Klass KD, Picker MD, Damgaard J, Van Noort S, Tojo K (2003). The taxonomy, genitalic morphology and phylogenetic relationships of southern african Mantophasmatodea (Insecta). Entomol. Abhandl. 61: 3–67.

Kjer KM, Carle FL, Litman J, Ware J (2006). A molecular phylogeny of Insecta. Arthropod System Phylogeny 64: 35–44.

Kjer KM, Honeycutt RL (2007). Site specific rates of mitochondrial genome and phylogeny of eutheria. BMC Evolutionary Biology 7, 8.

Kristensen NP (1975). The phylogeny of hexapod “orders.” A critical review of recent accounts. Zeitschr Zool Syst Evolsforsch 13: 1–44.

Kristensen NP (1981). Phylogeny of insect orders. Annu Rev Entomol 26: 135–57.

Kristensen NP (1991). Phylogeny of extant hexapods. In The Insects of Australia, ed. CSIRO, 1:125–40. Ithaca, NY: Cornell Univ. Press

Kristensen NP (1995). Forty years' insect phylogenetic systematics. Zool Beitr NF 36: 83–124.

Kristensen NP (1997). The groundplan and basal diversification of the hexapods. In Arthropod Relationships, ed. RA Fortey, RH Thomas. Syst Assoc Spec Vol Ser 55: 282–93.

Kuhner MK, Felsenstein J (1994). A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. Mol Biol Evol 11: 459–68.

Kukalova-Peck J (1991). Fossil history and the evolution of hexapod structures. The Insects of Australia, A Textbook for Students and Research Workers, 2nd ed., Melbourne University Press, Paris, pp. 141–179.

Kurabayashi A, Ueshima R (2000). Complete sequence of the mitochondrial DNA of the primitive opisthobranch gastropod *Pupa strigosa*: systematic implication of the genome organization. Mol Biol Evol 17: 266–277.

Labandeira C (2006). Silurian to Triassic plant and hexapod clades and their associations: new data, a review, and interpretations. Arthropod Syst Phylogeny 64: 53–94.

Ladoukakis ED, Zouros E (2001a). Direct evidence for homologous recombination in mussel (*Mytilus galloprovincialis*) mitochondrial DNA. Mol Biol Evol 18: 1168-75.

Ladoukakis ED, Zouros E (2001b). Recombination in animal mitochondrial DNA: evidence from published sequences. Mol Biol Evol 18: 2127-31.

Laudokakis ED, Saavedra C, Margoulas A, Zouros E (2002). Mitochondrial DNA variation in a species with two mitochondrial genomes: the case of *Mytilus galloprovincialis* from the Atlantic, the Mediterranean and the Black Sea. Mol Ecol 11: 755-769.

Lavrov DV, Brown WM (2001a). *Trichinella spiralis* mtDNA: a nematode mitochondrial genome

that encodes a putative ATP8 and normally structured tRNAS and has a gene arrangement relatable to those of coelomate metazoans. Genetics 157: 621–637.

Lavrov DV, Brown WM, Boore JL (2002). Complete mtDNA Sequences of Two Millipedes Suggest a New Model for Mitochondrial Gene Rearrangements: Duplication and Nonrandom Loss. Mol Biol Evol 19(2): 163–169.

Lenaz G (1998). Role of mitochondria in oxidative stress and ageing. Biochim Biophys Acta 1366: 53-67.

Levings CS III, Vasil IK (1995). The Molecular Biology of Plant Mitochondria. Kluwer Academic Publishers, Boston, MA.

Little CTS, Vrijenhoek RC (2003). Are hydrothermal vent animals living fossils? Trends ecol. Evol. 18: 582–588.

Liu H, Beckenbach AT (1992). Evolution of the mitochondrial oxidase II gene among 10 orders of insects. Mol Phyl Evol 1:41–52.

Liu HPMJ, Wu SK (1996). Paternal mitochondrial DNA differentiation far exceeds maternal DNA and allozyme differentiation in the freshwater mussel, *Anodonta grandis grandis*. Evolution 50.

Lockhart PJ, Larkum AWD, Steel MA, Waddell PJ, Penny D (1996). Evolution of chlorophyll and bacteriochlorophyll: the problem of invariant sites in sequence analysis. Proc Natl Acad Sci USA 93: 1930–34.

Lockhart P, Novis P, Milligan BG, Riden J, Rambaut A, Larkum T (2006). Heterotachy and tree building: a case study with plastids and eubacteria. Mol Biol Evol 23: 40–45.

Lockhart P, Steel M (2005). A tale of two processes Syst Biol 54: 948–51.

Longo FJ, Dornfield EJ (1967). The fine structure of spermatid differentiation in the mussel, Mytilus edulis. J Ultrastruct Res 20: 462-480.

Lowe TM, Eddy SR (1997). tRNAscan-SE: a program for improved detection of transfer RNA

genes in genomic sequence. Nucleic Acids Res 25: 955–964.

Maddison W, Maddison D (2003). MacClade ver 4.06. Sinauer Associates. Sunderland.

McBride HM, Neuspiel M, Wasiak S (2006). Mitochondria: more than just a powerhouse. Curr Biol 16: R551-60.

Macey JR, Larson A, Ananjeva NB, Fang Z, Papenfuss TJ, Macey RJ (1997). Two novel gene orders and the role of light-strand replication in rearrangement of the vertebrate mitochondrial genome. Mol Biol Evol 14: 91–104.

Macey JR, Schulte II JA, Larson A, Papenfuss TJ (1998). Tandem duplication via light-strand synthesis may provide a precursor for mitochondrial genomic rearrangement. Mol Biol Evol 15: 71–75.

Madsen CS, Ghivizzani SC, Hauswirth WW (1993). In vivo and in vitro evidence for slipped mispairing in mammalian mitochondria. Proc Natl Acad Sci USA 90: 7671–7675.

Medina M, Collins AG, Takaoka TL, Kuehl JV, Boore JL (2006). Naked corals: skeleton loss in Scleractinia. Proc Natl Acad Sci USA 103: 9096–9100.

Milbury CA, Gaffney PM (2005). Complete mitochondrial DNA sequence of the eastern oyster *Crassostrea virginica*. Mar Biotechnol (NY) 7: 697–712.

Mizi A, Zouros E, Moschonas N, Rodakis GC (2005). The complete maternal and paternal mitochondrial genomes of the Mediterranean mussel *Mytilus galloprovincialis*: implications for the doubly uniparental inheritance mode of mtDNA. Mol Biol Evol 22: 952–967.

Moritz C, Brown WM (1987). Tandem duplications in animal mitochondrial DNAs: variation in incidence and gene content among lizards. Proc Natl Acad Sci USA 84: 7183– 7187.

Morse DE, Duncan H, Hooker N, Morse A (1977). Hydrogen peroxide induces spawning in molluscs, with activation of prostaglandin endoperoxide synthtase. Science 196: 298-300.

Moss T, Stefanovsky VY (1995). Promotion and regulation of ribosomal transcription in eukaryotes

by RNA polymerase I. Prog Nucleic Acid Res Mol Biol 50: 25-66.

Mossel E, Steel M (2004). A phase transition for a random cluster model on phylogenetic trees. Math Biosci 187: 189–203.

Mossel E, Steel M (2005). How much can evolved characters tell us about the tree that generated them? In Mathematics of Evolution and Phylogeny, ed. O Gascuel, pp. 384–412. Oxford: Oxford Univ. Press.

Nardi F, Carapelli A, Fanciulli PP, Dallai R, Frati F (2001). The complete mitochondrial DNA sequence of the basal hexapod *Tetrodontophora bielanensis*: evidence for heteroplasmy and tRNA translocations. Mol Biol Evol 18: 1293–1304.

Nardi F, Spinsanti G, Boore JL, Carapelli A, Dallai R, Frati F (2003). Hexapod origins: monophyletic or paraphyletic? Science 299: 1887–1889.

Noguchi Y, Endo K, Tajima F, Ueshima R (2000). The mitochondrial genome of the brachiopod *Laqueus rubellus*. Genetics 155: 245–259.

Obata M, Kamiya C, Kawamura K, Komaru A (2006). Sperm mitochondrial DNA transmission to both male and female offspring in the blue mussel *Mytilus galloprovincialis*. Dev Growth Differ 48: 253-61.

Obata M, Komaru A. (2005) Specific location of sperm mitochondria in *mussel Mytilus galloprovincialis* zygotes stained by MitoTracker. Dev Growth Differ 47: 255-63.

Obata M, Shimizu M, Sano N, Komaru A (2008). Maternal Inheritance of Mitochondrial DNA (mtDNA) in the Pacific oyster (*Crassostrea gigas*): a Preliminary Study Using mtDNA Sequence Analysis with Evidence of Random Distribution of MitoTracker-Stained Sperm Mitochondria in Fertilized Eggs. Zoolog Sci 25: 248-54.

Ohta T (1992): The Nearly Neutral Theory Of Molecular Evolution. Ann Rev Ecol Sys 23

Ohtsuki T, Sato A, Watanabe Y, Watanabe K (2002). A unique serine-specific elongation factor Tu found in nematode mitochondria. Nat Struct Biol 9: 669–673.

Ojala D, Montoya J, Attardi G (1981). tRNA punctuation model of RNA processing in human mitochondria. Nature 290: 470–474.

Okimoto RH, Chamberlin M, Macfarlane JL, Wolstenholme DR (1991). Repeated sequence sets in mitochondrial DNA molecules of root knot nematodes (*Meloidogyne*): nucleotide sequences, genome location and potential for host race identification. Nucleic Acids Res 19: 1619–1626.

Okimoto R, Macfarlane JL, Clary DO, Wolstenholme DR (1992). The mitochondrial genomes of two nematodes, *Caenorhabditis elegans* and *Ascaris suum*. Genetics 130: 471–498.

Okimoto R, Macfarlane JL, Wolstenholme DR (1994). The mitochondrial ribosomal RNA genes of the nematodes *Caenorhabditis elegans* and *Ascaris suum*: consensus secondary-structure models and conserved nucleotide sets for phylogenetic analysis. J Mol Evol 39: 598–613.

Okimoto R, Wolstenholme DR (1990). A set of tRNAs that lack either the T psi C arm or the dihydrouridine arm: towards a minimal tRNA adaptor. EMBO J 9: 3405–3411.

Palumbi S, Martin A, Romano S, McMillan WO, Stice L, Grabowski G (1991): The simple fools guide to PCR Honolulu, Special publication Department of Zoology and Kewalo Marine Laboratory, University of Hawaii.

Papakonstantinou T, Law RH, Nagley P, Devenish RJ (1996a). Non-functional variants of yeast mitochondrial ATP synthase subunit 8 that assemble into the complex. Biochem Mol Biol Int 39: 253–260.

Papakonstantinou T, Law RH, Nesbitt WS, Nagley P, Devenish RJ (1996b). Molecular genetic analysis of the central hydrophobic domain of subunit 8 of yeast mitochondrial ATP synthase. Curr Genet 30: 12–18.

Papetti C, Lio P, Ruber L, Patarnello T, Zardoya R (2007). Antarctic fish mitochondrial genomes lack ND6 gene. J Mol Evol 65: 519–528.

Papillon D, Perez Y, Caubit X, Le Parco Y (2004). Identification of chaetognaths as protostomes is supported by the analysis of their mitochondrial genome. Mol Biol Evol 21: 2122–2129.

Paquin B, Laforest MJ, Forget L, Roewer I, Wang Z, Longcore J, Lang BF (1997). The fungal mitochondrial genomes and their gene expression. Curr. Genet. 31: 380-395.

Passamonti M (2007). An unusual case of gender-associated mitochondrial DNA heteroplasmy: the mytilid *Musculista senhousia* (Mollusca Bivalvia). BMC Evol Biol 7Suppl 2: S7.

Passamonti M, Boore JL, Scali V (2003). Molecular evolution and recombination in gender-associated mitochondrial DNAs of the Manila clam *Tapes philippinarum*. Genetics 164: 603-11.

Passamonti M, Ghiselli F (2009). Doubly uniparental inheritance: two mitochondrial genomes, one precious model for organelle DNA inheritance and evolution. DNA Cell Biol 28 (2):79-89.

Passamonti M, Scali V (2001). Gender-associated mitochondrial DNA heteroplasmy in the venerid clam *Tapes philippinarum* (Mollusca: Bi-valvia). Curr Genet 39: 117–124.

Peleg O, Kirzhner V, Trifonov E, Bolshoy A (2004). Overlapping messages and survivability. J Mol Evol 59: 520–527.

Perna NT, Kocher TD (1995). Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. J Mol Evol 41: 353–358.

Philippe H, Adouette A (1996). What can phylogenetic patterns tell us about the evolutionary processes generating biodiversity? In Aspects of the Genesis and Maintenance of Biological Diversity, ed. M Hochberg, J Clobert, R Barbault, pp. 41–59. Oxford: Oxford Univ. Press.

Philippe H, Laurent J (1998). How good are deep phylogenetic trees? Curr Opin Genet Dev 8: 616–23.

Phillips MJ, Delsuc F, Penny D (2004). Genome-scale phylogeny and the detection of systematic biases. Mol Biol Evol 21: 1455–58.

Posada D, Crandall KA (1998). ModelTest: testing the best-Wt model of nucleotide substitution. Bioinformatics 14: 817–818.

Quesada H et al. (2003). Heteroplasmy suggests paternal cotransmission of multiple genomes and pervasive reversion of maternally into paternally transmitted genomes of mussel (*Mytilus*) mitochondrial DNA. J Mol Evol 57 :S138–S147

Quesada H, Warren M, Skibinski DO (1998). Nonneutral evolution and differential mutation rate of gender-associated mitochondrial DNA lineages in the marine mussel *Mytilus*. Genetics 149: 1511-26.

Rand DM, Kann LM (1998). Mutation and selection at silent and replacement sites in the evolution of animal mitochondrial DNA. Genetica 102/103: 393–407.

Rawlings TA, Collins TM, Bieler R (2001). A major mitochondrial gene rearrangement among closely related species. Mol Biol Evol 18: 1604-1609.

Rawson PD (2005). Nonhomologous recombination between the large unassigned region of the male and female mitochondrial genomes in the mussel, *Mytilus trossulus*. J Mol Evol 61: 717–732

Rawson PD, Burton RS (2006). Molecular evolution at the cytochrome oxidase subunit 2 gene among divergent populations of the intertidal copepod, *Tigriopus californicus*. J Mol Evol 62: 753–764

Rawson PD, Hilbish TJ (1995). Evolutionary relationships among the male and female mitochondrial DNA lineages in the *Mytilus edulis* species complex. Mol Biol Evol 12: 893–901.

Rawson PD, Secor CL, Hilbish TJ (1996). The effects of natural hybridization on the regulation of doubly uniparental mtDNA inheritance in blue mussels (*Mytilus* spp.). Genetics 144: 241-8.

Reeder RH (1999). Regulation of RNA polymerase I transcription in yeast and vertebrates. Prog Nucleic Acid Res Mol Biol 62: 293-327.

Reese MG (2001): Application of a time-delay neural network to promoter annotation in the Drosophila melanogaster genome. Comput Chem 26: 51-56.

Roe BA, Ma DP, Wilson RK, Wong JF (1985). The complete nucleotide sequence of *the Xenopus laevis* mitochondrial genome. J Biol Chem 260: 9759–9774.

Ronquist F, Huelsenbeck JP (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19: 1572–1574.

Rokas A, Holland PWH (2000). Rare genomic changes as a tool for phylogenetics. Trends Ecol Evol 15: 454–59.

Rokas A, Ladoukakis E, Zouros E (2003). Animal mitochondrial DNA recombination revisited. Trends in Ecology and Evolution. 18.

Rokas A, Kruger D, Carroll SB (2005). Animal evolution and the molecular signature of radiations compressed in time. Science 310: 1933–38.

Rozas J, Sánchez-DelBarrio JC, Messegyer X, Rozas R (2003): DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19: 2496-2497.

Saavedra C, Reyero MI, Zouros E (1997). Male-dependent doubly uniparental inheritance of mitochondrial DNA and female-dependent sex-ratio in the mussel *Mytilus galloprovincialis*. Genetics. 145: 1073-82.

Saavedra C, Stewart DT, Stanwood RR, Zouros E (1996). Species-specific segregation of gender-associated mitochondrial DNA types in an area where two mussel species (*Mytilus edulis* and *M. trossulus*) hybridize. Genetics. 143:1359-67.

Sakurai M, Ohtsuki T, Watanabe Y, Watanabe K (2001). Requirement of modified residue m1A9 for EF-Tu binding to nematode mitochondrial tRNA lacking the T arm. Nucleic Acids Res Suppl 1: 237–238.

Saitoh K, Sado T, Mayden RL, Hanzawa N, Nakamura K, Nishida M, Miya M (2006). Mitogenomic evolution and interrelationships of the Cypriniformes (Actinopterygii: Ostariophysi): the first evidence toward resolution of higher-level relationships of the world's largest freshwater fish clade based on 59 whole mitogenome sequences. J Mol Evol 63: 826–841.

Sakurai M, Watanabe Y, Watanabe K, Ohtsuki T (2006). A protein extension to shorten RNA: elongated elongation factor-Tu recognizes the D-arm of T-armless tRNAs in nematode

mitochondria. Biochem J 399: 249–256.

Schapira AH (2006). Mitochondrial disease. Lancet 368: 70-82.

Schneider A, Ebert D (2004). Covariation of mitochondrial genome size with gene lengths: evidence for gene length reduction during mitochondrial evolution. Journal of Molecular Evolution 59: 90–96.

Serb JM, Lydeard C (2003). Complete mtDNA sequence of the North American freshwater mussel, *Lampsilis ornata* (Unionidae): an examination of the evolution and phylogenetic utility of mitochondrial genome organization in Bivalvia (Mollusca). Mol Biol Evol 20: 1854–1866.

Shao RF, Barker SC (2003). The highly rearranged mitochondrial genome of the plague thrips, *Thrips imaginis* (Insecta: thysanoptera): convergence of two novel gene boundaries and an extraordinary arrangement of rRNA genes. Mol Biol Evol 20: 362–370.

Shao R, Campbell NJH, Barker SC (2001). Numerous gene rearrangements in the mitochondrial genome of the wallaby louse, *Heterodoxus macropus* (Phthiraptera). Mol Biol Evol 18: 858–865.

Shao R, Mitani H, Barker SC, Takahashi M, Fukunaga M (2005b). Novel mitochondrial gene content and gene arrangement indicate illegitimate inter-mtDNA recombination in the chigger mite, *Leptotrombidium pallidum*. J Mol Evol 60: 764–773.

Sharov AG (1968). Phylogeny of the Orthopteroidea. Akademiya Nauk SSSR Trudy Paleontologicheskogo Instituta 118: 1–216.

Skibinski DO, Gallagher C, Beynon CM (1994a). Mitochondrial DNA inheritance. Nature. 368: 817-8.

Skibinski DO, Gallagher C, Beynon CM. (1994b) Sex-limited mitochondrial DNA transmission in the marine mussel *Mytilus* edulis. Genetics. 138:801-9.

Skibinski DOF, Gallagher C, Quesada U (1999). On the roles of selection, mutation and drift in the evolution of mitochondrial DNA diversity in British *Mytilus edulis* (Mytilidae; Mollusca)

populations. Biol J Linn Soc 68: 195-213.

Signorovitch AY, Buss LW, Dellaporta SL (2007). Comparative genomics of large mitochondria in placozoans. PLoS Genet 3: e13.

Simon C, Buckley TR, Frati F, Stewart JB, Beckenbach AT (2006). Incorporating molecular evolution into phylogenetic analysis, and a new compilation of conserved polymerase chain reaction primers for animal mitochondrial DNA. Annu Rev Ecol Evol Syst 37: 545–79.

Sinniger F, Chevaldonne P, Pawlowski J (2007). Mitochondrial genome of *Savalia savaglia* (Cnidaria, Hexacorallia) and early metazoan phylogeny. J Mol Evol 64: 196–203.

Skelton PW, Bentos MJ (1993). Mollusca: *Rostroconchia*, Scaphapoda and Bivalvia. Pp. 237–262 in M. J. Benton,ed. The fossil record 2. Chapman and Hall, London.

Spruyt N, Delarbre C, Gachelin G, Laudet V (1998). Complete sequence of the amphioxus (*Branchiostoma lanceolatum*) mitochondrial genome: relations to vertebrates. Nucleic Acids Res 26: 3279–3285.

Stechmann A, Schlegel M (1999). Analysis of the complete mitochondrial DNA sequence of the brachiopod *Terebratulina retusa* places Brachiopoda within the protostomes. Proc R Soc Lond B Biol Sci 266: 2043–2052.

Steel M, Huson D, Lockhart P (2000). Invariable site models and their use in phylogeny reconstruction. Syst Biol 49: 225–32.

Steinauer ML, Nickol BB, Broughton R, Orti G (2005). First sequenced mitochondrial genome from the phylum Acanthocephala (*Leptorhynchoides thecatus*) and its phylogenetic position within Metazoa. J Mol Evol 60: 706–715.

Stewart JB, Beckenbach AT (2003). Phylogenetic and genomic analysis of the complete mitochondrial DNA sequence of the spotted asparagus beetle *Crioceris duodecimpunctata*. Mol Phyl Evol 26: 513–526.

Stewart DT, Kenchington ER, Singh RK, Zouros E (1996). Degree of selective constraint as an explanation of the different rates of evolution of gender-specific mitochondrial DNA lineages in the mussel *Mytilus*. Genetics 143: 1349-1357.

Stewart DT, Saavedra C, Stanwood RR, Ball AO, Zouros E. (1995). Male and female mitochondrial DNA lineages in the blue mussel (*Mytilus edulis*) species group. Mol Biol Evol 12: 735-47.

Sueoka N (1962). On the genetic basis of variation and heterogeneity of DNA base composition. Proc Natl Acad Sci USA 48: 582–592.

Suga K, Mark Welch DB, Tanaka Y, Sakakura Y, Hagiwara A (2008). Two circular chromosomes of unequal copy number make up the mitochondrial genome of the rotifer *Brachionus plicatilis*. Mol Biol Evol 7: 7.

Sutherland BW, Stewart DT, Kenchington E, Kouros E (1998). The fate of paternal mitochondrial DNA in developing female mussels, *Mytilus edulis*: implications for the mechanism of doubly uniparental inheritance of mitochondrial DNA. Genetics 148: 341–347.

Sutovsky P, Moreno RD, Ramalho-Santos J, Dominko T, Simerly C, Schatten G (1999). Ubiquitin tag for sperm mitochondria. Nature 402: 371-2.

Swofford DL (2003): PAUP*. In Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4 Sunderland Massachusetts, Sinauer Associates.

Tang S, Hyman BC (2007). Mitochondrial genome haplotype hypervariation within the isopod parasitic nematode *Thaumamermis cosgrovei*. Genetics 176: 1139–1150.

Tamura K, Dudley J, Nei M, Kumar S (2007): MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24: 1596-1599.

Tateno Y, Takezaki N, Nei M (1994). Relative efficiencies of the maximum-likelihood, neighbor-joining and maximum-parsimony methods when substitution rate varies with site. Mol Biol Evol 11: 261–77.

Terrett JA, Miles S, Thomas RH (1996). Complete DNA sequence of the mitochondrial genome of

*Cepaea nemoralis* (Gastropoda: Pulmonata). J Mol Evol 42: 160–168.

Terry MD, Whiting MF (2005). Mantophasmatodea and phylogeny of the lower neopterous insects. Cladistics 21: 240–257.

Theologidis I, Fodelianakis S, Gaspar MB, Zouros E (2008). Doubly uniparental inheritance (DUI) of mitochondrial DNA in *Donax trunculus* (Bivalvia: Donacidae) and the problem of its sporadic detection in Bivalvia. Evol Int J Org Evolut. 62: 959-70.

Thompson JD, Higgins DG, Gibson TJ (1994). Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specifc gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673–4680.

Tilgner EH, Kiselyova TG, McHugh JV (1999). A morphological study of *Timema cristinae* Vickery with implications for the phylogenetics of Phasmida. Dtsche. Entomol. Z. 46, 149–162.

Tomita K, Yokobori S, Oshima T, Ueda T, Watanabe K (2002). The cephalopod *Loligo bleekeri* mitochondrial genome: multiplied noncoding regions and transposition of tRNA genes. J Mol Evol 54: 486–500.

Valles Y, Halanych KM, Boore JL (2008). Group II introns break new boundaries: presence in a Bilaterian's genome. PLoS ONE 3: e1488.

Valverde JR, Marco R, Garesse R (1994). A conserved heptamer motif for ribosomal RNA transcription termination in animal mitochondria. Proc Natl Acad Sci USA 91: 5368–5371.

Venetis C, Theologidis I, Zouros E, Rodakis GC (2006). No evidence for presence of maternal mitochondrial DNA in the sperm of *Mytilus galloprovincialis* males. Proc Biol Sci 273: 2483–2489.

Verdonk NH, Van Den Biggelaar JAM (1983). Early development and the formation of the germ layers. In The Mollusca: development Volume 3. Edited by: Verdonk NH, Van Den Biggelaar JAM, Tompa AS. New York, Academic Press: 91-122.

Wang X (2001). The expanding role of mitochondria in apoptosis. Genes Dev 15: 2922-33.

Wellman CH, Osterff PL, Mohiuddin U (2003). Fragments of the earliest land plants. Nature 425: 282–285.

Werren JH BL (1998). Sex determination, sex ratios and genetic conflict. Annu Rev Ecol Syst 29.

Wheeler WC, Whiting MF, Wheeler QD, Carpenter JM (2001). The phylogeny of the extant hexapod orders. Cladistics 17: 113–169.

Whiting MF (2002). Mecoptera is paraphyletic: multiple genes and phylogeny of Mecoptera and Siphonaptera. Zool Scr 31: 93–104.

Whiting MF, Bradler S, Maxwell T (2003). Loss and recovery of wings in stick insects. Nature 421: 264-267.

Whiting MF, Carpenter JM, Wheeler QD, Wheeler WC (1997). The Strepsiptera problem: phylogeny of the holometabolous insect orders inferred from 18S and 28S ribosomal DNA sequences and morphology. Syst Biol 46: 1–68.

Wolstenholme DR (1992). Animal mitochondrial DNA: structure and evolution. Int Rev Cytol 141: 173–216.

Yaffe MP (1999). Dynamic mitochondria. Nat Cell Biol 1: E149-50.

Yamazaki N, Ueshima R, Terrett JA (1997). Evolution of pulmonate gastropod mitochondrial genomes: comparisons of gene organizations of *Euhadra*, *Cepaea* and *Albinaria* and implications of unusual tRNA secondary structures. Genetics 145:749–758.

Yang Z (1995). Evaluation of several methods for estimating phylogenetic trees when substitution rates differ over nucleotide sites. J Mol Evol 40: 689–97.

Yang Z (1996). Among-site rate variation and its impact on phylogenetic analysis. Trends Ecol Evol 11: 367–72.

Yen MR, Tseng YH, Nguyen EH, Wu LF, Saier Jr MH (2002). Sequence and phylogenetic analyses

of the twin-arginine targeting (*Tat*) protein export system. Arch Microbiol 177: 441–450.

Yokobori S, Fukuda N, Nakamura M, Aoyama T, Oshima T (2004). Long-term conservation of six duplicated structural genes in cephalopod mitochondrial genomes. Mol Biol Evol 21: 2034–2046.

Yokobori S, Oshima T, Wada H (2005). Complete nucleotide sequence of the mitochondrial genome of *Doliolum nationalis* with implications for evolution of urochordates. Mol Phyl Evol 34: 273–283.

Yokobori S, Ueda T, Feldmaier-Fuchs G, Paabo S, Ueshima R, Kondow A et al. (1999). Complete DNA sequence of the mitochondrial genome of the ascidian *Halocynthia roretzi* (Chordata, Urochordata). Genetics 153: 1851–1862.

Yokobori S, Watanabe Y, Oshima T (2003). Mitochondrial genome of *Ciona savignyi* (Urochordata, Ascidiacea, Enterogona): comparison of gene arrangement and tRNA genes with *Halocynthia roretzi* mitochondrial genome. J Mol Evol 57: 574–587.

Zbawicka M, Burzynski A, Wenne R (2007). Complete sequences of mitochondrial genomes from the Baltic mussel *Mytilus trossulus*. Gene 406: 191–198.

Zhang J, Zhou, Gay J, Song D, Zhouk K (2008). The complete mitochondrial genome of *Parafronurus youi* (Insecta: Ephemeroptera) and phylogenetic position of the Ephemeroptera. Gene 15: 18-24.

Zouros E, Ball AO, Saavedra C, Freeman KR (1994a). Mitochondrial DNA inheritance. Nature. 368:818.

Zouros E, Freeman KR, Ball AO, Pogson GH (1992). Direct evidence for extensive paternal mitochondrial DNA inheritance in the marine mussel *Mytilus*. Nature 359: 412–414.

Zouros E, Oberhauser Ball A, Saavedra C, Freeman KR (1994b). An unusual type of mitochondrial DNA inheritance in the blue mussel Mytilus. Proc Natl Acad Sci USA 91: 7463-7467.

Zucker M (2003). Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31: 3406-3415.

## 6. TABLES

Table1: Lenght, base composition and Sequence divergence of M and F gene and Unassigned Regions in *Musculista senhousia*

| | | | | Base Composition (%) | | | | Divergence (SE) | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Gene/Region | F/M type | Lenght | T | C | A | G | pD | $K_s$ | $K_a$ |
| Noncoding | UR1-26/LUR | M | 4556 bp | 37.9 | 11.4 | 31.4 | 19.3 | NA | NA | NA |
| | UR1-28/LUR | F | 7007 bp | 38.0 | 10.5 | 30.7 | 20.8 | NA | NA | NA |
| rRNA Genes | rrnaL | M | 1682 bp | 37.3 | 12.6 | 30.8 | 19.3 | 0.312 (0.013) | NA | NA |
| | | F | 1125 bp | 35.8 | 13.4 | 30.4 | 20.4 | | | |
| | rrnaS | M | 1087 bp | 36.3 | 12.0 | 31.6 | 20.1 | 0.093 (0.009) | NA | NA |
| | | F | 819 bp | 37.2 | 11.0 | 32.1 | 19.7 | | | |
| | All rRNA genes | M | 2769 bp | 36.9 | 12.4 | 31.1 | 19.6 | 0.690 (0.010) | NA | NA |
| | | F | 1944 bp | 36.4 | 12.4 | 31.1 | 20.1 | | | |
| Protein genes | atp6 | M | 714 bp | 42.2 | 12.9 | 23.8 | 21.1 | 0.258 (0.017) | 0.894 (0.116) | 0.156 (0.021) |
| | | F | 714 bp | 43.8 | 12.7 | 23.5 | 19.9 | | | |
| | cox1 | M | 1584 bp | 38.3 | 15.9 | 24.7 | 21.1 | 0.180 (0.009) | 0.838 (0.073) | 0.042 (0.006) |
| | | F | 1584 bp | 40.0 | 14.4 | 24.4 | 21.3 | | | |
| | cox2 | M | 690 bp | 36.7 | 15.2 | 26.7 | 21.4 | 0.264 (0.016) | 0.887 (0.124) | 0.178 (0.024) |
| | | F | 660 bp | 37.4 | 14.5 | 27.3 | 20.8 | | | |
| | cox2b | M | 813 bp | 35.9 | 14.1 | 28.7 | 21.3 | 0.267* (0.016) | 0.653* (0.088) | 0.223* (0.030) |
| | | F | NA | | NA | | | NA | NA | NA |
| | cox3 | M | 855 bp | 42.0 | 13.1 | 23.3 | 21.6 | 0.220 (0.012) | 0.811 (0.094) | 0.107 (0.016) |
| | | F | 855 bp | 43.4 | 12.9 | 20.9 | 22.8 | | | |
| | cob | M | 1197 bp | 40.6 | 13.9 | 25.2 | 20.3 | 0.106 (0.009) | 0.346 (0.037) | 0.034 (0.007) |
| | | F | 1197 bp | 40.4 | 13.6 | 24.9 | 21.1 | | | |
| | nad1 | M | 996 bp | 39.8 | 12.2 | 26.0 | 22.0 | 0.228 (0.012) | 0.673 (0.069) | 0.146 (0.017) |
| | | F | 993 bp | 41.3 | 11.5 | 24.4 | 23.2 | | | |
| | nad2 | M | 945 bp | 44.9 | 10.8 | 24.4 | 19.9 | 0.302 (0.013) | 0.843 (0.086) | 0.244 (0.025) |
| | | F | 945 bp | 44.1 | 10.9 | 22.4 | 22.5 | | | |
| | nad3 | M | 375 bp | 44.3 | 14.1 | 21.3 | 20.3 | 0.261 (0.023) | 1.005 (0.196) | 0.147 (0.027) |
| | | F | 390 bp | 45.6 | 12.6 | 21.0 | 20.8 | | | |
| | nad4 | M | 1329 bp | 41.4 | 11.5 | 23.6 | 23.5 | 0.273 (0.011) | 0.931 (0.084) | 0.175 (0.017) |
| | | F | 1320 bp | 39.9 | 11.9 | 24.3 | 23.9 | | | |
| | nad4L | M | 216 bp | 43.5 | 8.8 | 24.5 | 23.1 | 0.199 (0.027) | 0.626 (0.143) | 0.107 (0.027) |
| | | F | 216 bp | 44.0 | 8.8 | 24.5 | 22.7 | | | |
| | nad5 | M | 1766 bp | 39.3 | 13.2 | 28.3 | 19.2 | 0.287 (0.010) | 0.839 (0.063) | 0.219 (0.017) |
| | | F | 1750 bp | 38.8 | 13.2 | 25.6 | 22.3 | | | |
| | nad6 | M | 624 bp | 43.8 | 11.4 | 25.6 | 19.2 | 0.284 (0.018) | 0.619 (0.082) | 0.268 (0.036) |
| | | F | 624 bp | 42.1 | 12.3 | 25.2 | 20.4 | | | |
| | All proteins | M | #11504 bp | 40.9 | 13.0 | 25.2 | 20.9 | #0.237 (0.003) | 0.750 (0.022) | 0.146 (0.005) |
| | | F | 11326 bp | 40.8 | 12.8 | 24.1 | 22.3 | | | |
| Complete | | M | 20612 bp | 39.3 | 12.7 | 27.7 | 20.3 | NA | NA | NA |
| | | F | 21557 bp | 39.3 | 12.0 | 27.2 | 21.4 | | | |

*: pD beetween male-*cox2* and *cox2b* genes

#: male-*cox2b* gene was excluded from the computation of overall pD

Table2: Organization of male *Musculista senhousia* mitochondrial genome.

| Position | Strand | Lenght | Gene and UR* | Anticodon | Start codon | Stop codon |
|---|---|---|---|---|---|---|
| 1-375 | H | 376 bp | *nad3* | | ATG | TAA |
| 376-434 | | 59 bp | UR-1 | | | |
| 435-500 | H | 66 bp | *trnY* | GTA | | |
| 501-534 | | 34 bp | UR-2 | | | |
| 535-598 | H | 64 bp | *trnH* | GTG | | |
| 599-619 | | 21 bp | UR-3 | | | |
| 620-687 | H | 68 bp | *trnI* | GAT | | |
| 688-751 | H | 64 bp | *trnN* | GTT | | |
| 752-3598 | H | 2847 bp | CR | | | |
| 3599-3667 | H | 69 bp | *trnE* | TTC | | |
| 3668-3708 | | 41 bp | UR-4 | | | |
| 3709-5292 | H | 1584 bp | *coxI* | | ATG | TAA |
| 5293-5852 | | 560 bp | UR-5 | | | |
| 5853-6665 | H | 813 bp | *coxIIb* | | ATG | TAA |
| 6666-6706 | | 41 bp | UR-6 | | | |
| 6707-7396 | H | 690 bp | *coxII* | | ATA | TAA |
| 7397-7612 | | 216 bp | UR-7 | | | |
| 7613-8326 | H | 714 bp | *atp6* | | ATG | TAA |
| 8327-8348 | | 22 bp | UR-8 | | | |
| 8349-8415 | H | 67 bp | *trnT* | TGT | | |
| 8392-9588 | H | 1197 bp | *Cob* | | ATA | TAA |
| 9589-9606 | | 18 bp | UR-9 | | | |
| 9607-9671 | H | 65 bp | *trnD* | GTC | | |
| 9672-9673 | | 2 bp | UR-10 | | | |
| 9674-9747 | H | 74 bp | *trnR* | TCG | | |
| 9748-9826 | | 79 bp | UR-11 | | | |
| 9827-9892 | H | 66 bp | *trnG* | TCC | | |
| 9893-9824 | L | 70 bp | *trnS* | GGA | | |
| 9893-10979 | | 1087 bp | *rrnaS* | | | |
| 10980-11417 | H | 438 bp | *nad6* | | ATA | TAA |
| 11418-11472 | | 55 bp | UR-12 | | | |
| 11473-12417 | H | 945 bp | *nad2* | | ATA | TAG |
| 12418-12444 | | 27 bp | UR-13 | | | |
| 12445-13299 | H | 855 bp | *cox3* | | ATG | TAA |
| 13299-13366 | H | 68 bp | *trnK* | TTT | | |
| 13367-13377 | | 11 bp | UR-14 | | | |
| 13378-13445 | H | 68 bp | *trnF* | GAA | | |
| 13446-13464 | | 19 bp | UR-15 | | | |
| 13465-13528 | H | 64 bp | *trnP* | TGG | | |
| 13529-13555 | | 27 bp | UR-16 | | | |
| 13556-13620 | H | 65 bp | *trnL2* | TAG | | |
| 13621-13626 | | 6 bp | UR-17 | | | |
| 13627-13694 | H | 68 bp | *trnC* | GCA | | |
| 13695-13738 | | 44 bp | UR-18 | | | |
| 13739-13803 | H | 65 bp | *trnL1* | TAA | | |
| 13804-13840 | | 37 bp | UR-19 | | | |
| 13841-14836 | H | 996 bp | *nad1* | | ATG | TAG |
| 14836-14898 | H | 63 bp | *trnM2* | TAT | | |
| 14899-14985 | | 87 bp | UR-20 | | | |
| 14986-15049 | H | 64 bp | *trnV* | TAC | | |
| 15050-15183 | | 134 bp | UR-21 | | | |
| 15184-15399 | H | 216 bp | *nad4L* | | ATA | TAA |
| 15400-15464 | | 65 bp | UR-22 | | | |
| 15465-17230 | H | 1766 bp | *nad5* | | ATA | T- |
| 17231-17294 | H | 65 bp | *trnA* | TGC | | |
| 17295-17338 | | 44 bp | UR-23 | | | |
| 17339-18667 | H | 1329 bp | *nad4* | | ATA | TAA |
| 18668-18710 | | 43 bp | UR-24 | | | |
| 18711-18776 | H | 66 bp | *trnW* | TCA | | |
| 18777-18781 | | 5 bp | UR-25 | | | |
| 18782-18848 | H | 67 bp | *trnQ* | TTG | | |
| 18849-18863 | | 15 bp | UR-26 | | | |
| 18864-18930 | H | 67 bp | *trnM1* | CAT | | |
| 18931-20612 | | 1682 bp | *rrnaL* | | | |

Note: *UR indicates Unassigned Regions

Table3: Organization of female *Musculista senhousia* mitochondrial genome.

| Position | Strand | Lenght | Gene and UR* | Anticodon | Start codon | Stop codon |
|---|---|---|---|---|---|---|
| 1-390 | H | 390 bp | *nad3* | | ATG | TAA |
| 391-625 | | 235 bp | UR-1 | | | |
| 626-691 | H | 66 bp | *trnY* | GTA | | |
| 692-1234 | | 542 bp | UR-2 | | | |
| 1235-1299 | H | 65 bp | *trnH* | GTG | | |
| 1300-1315 | | 16 bp | UR-3 | | | |
| 1316-1381 | H | 66 bp | *trnI* | GAT | | |
| 1382-1392 | | 11 bp | UR-4 | | | |
| 1393-1456 | H | 64 bp | *trnN* | GTT | | |
| 1457-1565 | | 109 bp | UR-5 | | | |
| 1566-1630 | H | 65 bp | *trnE* | TTC | | |
| 1631-6152 | | 4567 bp | CR | | | |
| 6153-7736 | H | 1540 bp | *cox1* | | ATG | TAA |
| 7737-8114 | | 378 bp | UR-6 | | | |
| 8115-8774 | H | 660 bp | *cox2* | | ATA | TAA |
| 8775-9051 | | 277 bp | UR-7 | | | |
| 9052-9765 | H | 714 bp | *atp6* | | ATG | TAG |
| 9766-9791 | | 26 bp | UR-8 | | | |
| 9792-9858 | H | 67 bp | *trnT* | TGT | | |
| 9835-11031 | H | 1197 bp | *cob* | | ATA | TAA |
| 11032-11049 | | 18 bp | UR-9 | | | |
| 11050-11114 | H | 65 bp | *trnD* | GTC | | |
| 11115-11123 | | 9 bp | UR-10 | | | |
| 11124-11189 | H | 66 bp | *trnR* | TCG | | |
| 11190-11269 | | 80 bp | UR-11 | | | |
| 11270-11335 | H | 66 bp | *trnG* | TCC | | |
| 11267-11336 | L | 70 bp | *trnS* | GGA | | |
| 11336-12154 | | 819 bp | *rrnaS* | | | |
| 12155-12778 | H | 624 bp | *nad6* | | ATG | TAA |
| 12779-12828 | | 50 bp | UR-12 | | | |
| 12829-13773 | H | 945 bp | *nad2* | | ATA | TAA |
| 13774-13855 | | 112 bp | UR-13 | | | |
| 13856-14710 | H | 855 bp | *cox3* | | ATG | TAA |
| 14711-14723 | | 13 bp | UR-14 | | | |
| 14724-14790 | H | 67 bp | *trnK* | TTT | | |
| 14791-14797 | | 7 bp | UR-15 | | | |
| 14798-14865 | H | 68 bp | *trnF* | GAA | | |
| 14866-14879 | | 14 bp | UR-16 | | | |
| 14880-14944 | H | 65 bp | *trnP* | TGG | | |
| 14945-14977 | | 33 bp | UR-17 | | | |
| 14978-15042 | H | 65 bp | *trnL2* | TAG | | |
| 15043-15048 | | 6 bp | UR-18 | | | |
| 15049-15112 | H | 64 bp | *trnC* | GCA | | |
| 15113-15159 | | 47 bp | UR-19 | | | |
| 15160-15223 | H | 64 bp | *trnL1* | TAA | | |
| 15224-15259 | | 36 bp | UR-20 | | | |
| 15260-16252 | H | 993 bp | *nad1* | | ATG | TAA |
| 16253-16385 | | 133 bp | UR-21 | | | |
| 16386-16448 | H | 63 bp | *trnM2* | TAT | | |
| 16449-16487 | | 39 bp | UR-22 | | | |
| 16488-16549 | H | 62 bp | *trnV* | TAC | | |
| 16550-16695 | | 146 bp | UR-23 | | | |
| 16696-16911 | H | 216 bp | *nad4L* | | ATA | TAA |
| 16912-16988 | | 77 bp | UR-24 | | | |
| 16989-18738 | H | 1750 bp | *nad5* | | ATA | T- |
| 18739-18804 | H | 66 bp | *trnA* | TGC | | |
| 18805-18843 | | 39 bp | UR-25 | | | |
| 18844-20163 | H | 1320 bp | *nad4* | | ATA | TAG |
| 20164-20213 | | 50 bp | UR-26 | | | |
| 20214-20280 | H | 67 bp | *trnW* | TCA | | |
| 20281-20285 | | 5 bp | UR-27 | | | |
| 20286-20353 | H | 68 bp | trnQ | TTG | | |
| 20354-20361 | | 7 bp | UR-28 | | | |
| 20362-20432 | H | 70 bp | *trnM1* | CAT | | |
| 20433-21557 | H | 1125 bp | *rrnaL* | | | |

Note: *UR indicates Unassigned Regions

Table4: genes, initiation and termination codons and number of amino acids in male and female *Musculista senhousia* genomes.

| **Protein gene** | **male start/stop cod.** | **female start/stop cod.** | **Maa** | **Faa** | **Distance(SE)** | **Ks** | **Ka** | **Ka/Ks** |
|---|---|---|---|---|---|---|---|---|
| *atp6* | ATG/TAA | ATG/TAG | 238 | 238 | 0.228 (0.017) | 0.894 (0.116) | 0.156 (0.021) | 0.17 |
| *cox1* | ATG/TAA | ATG/TAA | 1584 | 1584 | 0.053 (0.010) | 0.838 (0.073) | 0.042 (0.006) | 0.05 |
| *cox2* | ATA/TAA | ATA/TAA | 230 | 220 | 0.251 (0.029) | 0.887 (0.124) | 0.178 (0.024) | 0.20 |
| *cox2b* | ATG/TAA | NA | 271 | NA | 0.279* (0.029) | 0.653* (0.088) | 0.223* (0.030) | *0.34 |
| *cox3* | ATG/TAA | ATG/TAA | 285 | 285 | 0.155 (0.021) | 0.811 (0.094) | 0.107 (0.016) | 0.13 |
| *cob* | ATA/TAA | ATA/TAA | 399 | 399 | 0.058 (0.012) | 0.346 (0.037) | 0.034 (0.007) | 0.09 |
| *nad1* | ATG/TAG | ATG/TAA | 332 | 331 | 0.218 (0.021) | 0.673 (0.069) | 0.146 (0.017) | 0.21 |
| *nad2* | ATA/TAG | ATA/TAA | 315 | 315 | 0.306 (0.025) | 0.843 (0.086) | 0.244 (0.025) | 0.29 |
| *nad3* | ATG/TAA | ATG/TAA | 125 | 130 | 0.202 (0.036) | 1.005 (0.196) | 0.147 (0.027) | 0.14 |
| *nad4* | ATA/TAA | ATA/TAG | 443 | 440 | 0.243 (0.019) | 0.931 (0.084) | 0.175 (0.017) | 0.18 |
| *nad4L* | ATA/TAA | ATA/TAA | 72 | 72 | 0.183 (0.049) | 0.626 (0.143) | 0.107 (0.027) | 0.17 |
| *nad5* | ATA/T- | ATA/T- | 588 | 583 | 0.278 (0.017) | 0.839 (0.063) | 0.219 (0.017) | 0.26 |
| *nad6* | ATA/TAA | ATG/TAA | 208 | 208 | 0.324 (0.018) | 0.619 (0.082) | 0.268 (0.036) | 0.43 |
| All proteins | | | | | | 0.750 (0.022) | 0.146 (0.005) | 0.19 |

Note.- **start/stop cod.** indicate initiation and termination codons in male and female respectively.
**T-** : **TAA** stop codon is completed by additino of 3' A residue sto the mRNA.
**Maa** and **Faa** indicate the number of amino acids in male and female respectively.
**Distance (SE)** indicate p-Distances at the amino acidic level.
**Ks** and **Ka** indicate the divergence of protein genes in synonymous (**Ks**) and non synonymous (**Ka**) sites respectively.
**Ka/Ks** indicates ratio's value between **Ka** and **Ks.**

Table5: Codon usage in male and female *Musculista senhousia* mitochondrial genome.

| aa | Codon | N | % | aa | Codon | N | % | aa | Codon | N | % | aa | Codon | N | % |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Phe** | **TTT** | **324** | **8.45** | **Ser** | **TCT** | **118** | **3.06** | **Tyr** | **TAT** | **119** | **3.09** | **Cys** | **TGT** | **79** | **2.05** |
| | | **301** | **7.82** | | | **105** | **5.88** | | | **117** | **3.04** | | | **77** | **2.02** |
| | **TTC** | **46** | **1.19** | | **TCC** | **16** | **0.41** | | **TAC** | **35** | **0.91** | | **TGC** | **15** | **0.39** |
| | | **41** | **1.06** | | | **10** | **0.26** | | | **40** | **1.04** | | | **12** | **0.31** |
| **Leu** | **TTA** | **264** | **6.86** | | **TCA** | **34** | **0.88** | **s.c.** | **TAA** | **9** | **0.23** | **Trp** | **TGA** | **57** | **1.48** |
| | | **253** | **6.57** | | | **35** | **0.91** | | | **9** | **0.23** | | | **50** | **1.3** |
| | **TTG** | **104** | **2.70** | | **TCG** | **4** | **0.10** | | **TAG** | **2** | **0.05** | | **TGG** | **43** | **1.11** |
| | | **112** | **2.91** | | | **14** | **0.36** | | | **3** | **0.07** | | | **45** | **1.17** |
| **Leu** | **CTT** | **88** | **2.28** | **Pro** | **CCT** | **89** | **2.13** | **His** | **CAT** | **46** | **1.19** | **Arg** | **CGT** | **39** | **1.01** |
| | | **83** | **2.15** | | | **93** | **2.41** | | | **55** | **1.43** | | | **34** | **0.88** |
| | **CTC** | **14** | **0.36** | | **CCC** | **11** | **0.28** | | **CAC** | **25** | **0.65** | | **CGC** | **12** | **0.31** |
| | | **20** | **0.52** | | | **10** | **0.26** | | | **17** | **0.44** | | | **6** | **0.15** |
| | **CTA** | **54** | **1.14** | | **CCA** | **18** | **0.46** | **Gln** | **CAA** | **35** | **0.91** | | **CGA** | **8** | **0.20** |
| | | **65** | **1.69** | | | **12** | **0.31** | | | **32** | **0.83** | | | **16** | **0.41** |
| | **CTG** | **23** | **0.59** | | **CCG** | **8** | **0.20** | | **CAG** | **16** | **0.41** | | **CGG** | **10** | **0.26** |
| | | **46** | **1.19** | | | **8** | **0.20** | | | **23** | **0.59** | | | **13** | **0.33** |
| **Ile** | **ATT** | **161** | **4.18** | **Thr** | **ACT** | **57** | **1.48** | **Asn** | **AAT** | **79** | **2.05** | **Ser** | **AGT** | **77** | **2.02** |
| | | **155** | **4.03** | | | **54** | **1.40** | | | **78** | **2.02** | | | **67** | **1.74** |
| | **ATC** | **43** | **1.11** | | **ACC** | **19** | **0.49** | | **AAC** | **51** | **1.32** | | **AGC** | **33** | **0.85** |
| | | **40** | **1.04** | | | **7** | **0.18** | | | **29** | **0.75** | | | **32** | **0.83** |
| **Met** | **ATA** | **142** | **3.69** | | **ACA** | **35** | **0.91** | **Lys** | **AAA** | **100** | **2.60** | **Arg** | **AGA** | **93** | **2.41** |
| | | **149** | **3.87** | | | **30** | **0.78** | | | **81** | **2.10** | | | **99** | **2.57** |
| | **ATG** | **76** | **2.05** | | **ACG** | **11** | **0.28** | | **AAG** | **29** | **0.75** | | **AGG** | **78** | **2.02** |
| | | **59** | **1.53** | | | **15** | **0.39** | | | **29** | **0.75** | | | **63** | **1.63** |
| **Val** | **GTT** | **178** | **4.62** | **Ala** | **GCT** | **83** | **2,15** | **Asp** | **GAT** | **61** | **1.58** | **Gly** | **GGT** | **100** | **2.60** |
| | | **192** | **4.99** | | | **85** | **2.21** | | | **56** | **1.45** | | | **102** | **2.65** |
| | **GTC** | **29** | **0.75** | | **GCC** | **23** | **0.59** | | **GAC** | **17** | **0.44** | | **GGC** | **27** | **0.70** |
| | | **26** | **0.67** | | | **15** | **0.39** | | | **17** | **0.44** | | | **39** | **1.01** |
| | **GTA** | **106** | **2.75** | | **GCA** | **41** | **1.06** | **Glu** | **GAA** | **52** | **1.35** | | **GGA** | **48** | **0.70** |
| | | **119** | **3.09** | | | **45** | **1.17** | | | **45** | **1.17** | | | **44** | **1.14** |
| | **GTG** | **81** | **2.10** | | **GCG** | **20** | **0.52** | | **GAG** | **38** | **0.98** | | **GGG** | **81** | **2.10** |
| | | **83** | **2.15** | | | **20** | **0.52** | | | **49** | **1.27** | | | **93** | **2.41** |

Note.- In each codon the upper value correspond to M and the lower to F genome. Codons that match the corresponding tRNA anticodon are **underlined**.
aa: amino acid.
N: number of occurence of the codon in the genome.

Table6: female *Musculista senhousia*, UR (Unassigned regions) and LUR (Large Unassigned Region): base compositions and A+T content.

| | | Base Composition (%) | | | | A+T% |
|---|---|---|---|---|---|---|
| | Lenght | T | C | A | G | |
| UR-1 | 235 bp | 39.6 | 8.5 | 34.0 | 17.9 | 73.6 |
| UR-2 | 543 bp | 37.8 | 9.8 | 31.3 | 21.2 | 69.1 |
| UR-3 | 16 bp | 25.0 | 12.5 | 31.3 | 31.3 | 56.3 |
| UR-4 | 11 bp | 18.2 | 27.3 | 36.4 | 18.2 | 54.6 |
| UR-5 | 109 bp | 36.7 | 13.8 | 29.4 | 20.2 | 66.1 |
| UR-6 | 378 bp | 43.1 | 8.7 | 28.6 | 19.6 | 71.7 |
| UR-7 | 277 bp | 43.7 | 11.9 | 26.4 | 18.1 | 70.1 |
| UR-8 | 26 bp | 61.5 | 11.5 | 19.2 | 7.7 | 80.7 |
| UR-9 | 19 bp | 42.1 | 15.8 | 42.1 | 0.0 | 84.2 |
| UR-10 | 9 bp | 44.4 | 11.1 | 11.1 | 33.3 | 55.5 |
| UR-11 | 77 bp | 35.1 | 16.9 | 28.6 | 19.5 | 63.7 |
| UR-12 | 50 bp | 46.0 | 6.0 | 32.0 | 16.0 | 78.0 |
| UR-13 | 82 bp | 32.9 | 15.9 | 35.4 | 15.9 | 68.3 |
| UR-14 | 13 bp | 53.8 | 7.7 | 23.1 | 15.4 | 76.9 |
| UR-15 | 7 bp | 28.6 | 0.0 | 71.4 | 0.0 | 100.0 |
| UR-16 | 14 bp | 42.9 | 0.0 | 21.4 | 35.7 | 64.3 |
| UR-17 | 33 bp | 21.2 | 9.1 | 42.4 | 27.3 | 63.6 |
| UR-18 | 6 bp | 16.7 | 0.0 | 66.7 | 16.7 | 83.4 |
| UR-19 | 47 bp | 34.0 | 6.4 | 42.6 | 17.0 | 76.6 |
| UR-20 | 36 bp | 41.7 | 5.6 | 36.1 | 16.7 | 77.8 |
| UR-21 | 133 bp | 34.6 | 10.5 | 39.8 | 15.0 | 74.4 |
| UR-22 | 39 bp | 35.9 | 10.3 | 43.6 | 10.3 | 79.5 |
| UR-23 | 146 bp | 46.6 | 6.8 | 26.0 | 20.5 | 72.6 |
| UR-24 | 77 bp | 44.2 | 10.4 | 27.3 | 18.2 | 71.5 |
| UR-25 | 39 bp | 41.0 | 2.6 | 33.3 | 23.1 | 74.3 |
| UR-26 | 50 bp | 50.0 | 10.0 | 26.0 | 14.0 | 76.0 |
| UR-27 | 5 bp | 40.0 | 0.0 | 40.0 | 20.0 | 80.0 |
| UR-28 | 8 bp | 37.5 | 12.5 | 37.5 | 12.5 | 75.0 |
| LUR | 4520 bp | 36.8 | 10.8 | 30.5 | 21.9 | 67.3 |
| UR 1-28/LUR | 7007 bp | 38.0 | 10.5 | 30.7 | 20.8 | 68.7 |

Table7: male *Musculista senhousia*, UR (Unassigned regions) and LUR (Large Unassigned Region): base compositions and A+T content.

| | | Base Composition (%) | | | | A+T% |
|---|---|---|---|---|---|---|
| | | T | C | A | G | |
| **UR-1** | **59 bp** | **47.5** | **8.5** | **35.6** | **8.5** | **83.1** |
| **UR-2** | **34 bp** | **32.4** | **2.9** | **44.1** | **20.6** | **76.5** |
| **UR-3** | **21 bp** | **33.3** | **19.0** | **33.4** | **14.3** | **66.7** |
| **UR-4** | **41 bp** | **39.0** | **9.8** | **24.4** | **26.8** | **63.4** |
| **UR-5** | **560 bp** | **41.1** | **11.3** | **29.3** | **18.4** | **70.4** |
| **UR-6** | **41 bp** | **41.5** | **7.3** | **26.8** | **24.4** | **68.3** |
| **UR-7** | **216 bp** | **40.7** | **14.8** | **27.8** | **16.7** | **68.5** |
| **UR-8** | **22 bp** | **13.6** | **22.7** | **50.0** | **13.6** | **63.6** |
| **UR-9** | **18 bp** | **33.3** | **16.7** | **44.4** | **5.6** | **77.7** |
| **UR-10** | **2 bp** | **0.0** | **0.0** | **100.0** | **0.0** | **/** |
| **UR-11** | **78 bp** | **41.0** | **12.8** | **28.2** | **17.9** | **69.2** |
| **UR-12** | **56 bp** | **44.6** | **10.7** | **32.1** | **12.5** | **76.7** |
| **UR-13** | **27 bp** | **40.7** | **7.4** | **44.4** | **7.4** | **85.1** |
| **UR-14** | **11 bp** | **18.2** | **18.2** | **63.6** | **0.0** | **81.8** |
| **UR-15** | **19 bp** | **31.6** | **15.8** | **36.8** | **15.8** | **68.4** |
| **UR-16** | **27 bp** | **44.4** | **0.0** | **40.7** | **14.8** | **85.1** |
| **UR-17** | **6 bp** | **33.3** | **0.0** | **50.0** | **16.7** | **83.3** |
| **UR-18** | **44 bp** | **36.4** | **13.6** | **29.5** | **20.5** | **65.9** |
| **UR-19** | **37 bp** | **27.0** | **16.2** | **45.9** | **10.8** | **72.9** |
| **UR-20** | **86 bp** | **45.3** | **12.8** | **24.4** | **17.4** | **79.7** |
| **UR-21** | **134 bp** | **46.3** | **5.2** | **28.4** | **20.1** | **74.7** |
| **UR-22** | **64 bp** | **45.3** | **10.9** | **29.7** | **14.1** | **75.0** |
| **UR-23** | **44 bp** | **29.5** | **13.6** | **26.0** | **20.5** | **55.5** |
| **UR-24** | **43 bp** | **51.2** | **9.3** | **27.9** | **11.6** | **79.1** |
| **UR-25** | **5 bp** | **40.0** | **20.0** | **40.0** | **0.0** | **80.0** |
| **UR-26** | **15 bp** | **26.7** | **0.0** | **53.0** | **20.0** | **79.7** |
| **LUR** | **2.846 bp** | **36.3** | **11.5** | **31.4** | **20.7** | **67.7** |
| **UR 1-26/LUR** | **4.556 bp** | **37.9** | **11.4** | **31.4** | **19.3** | **69.3** |

Table8: Organization of *Bacillus rossius* mitochondrial genome.

| Position | Gene | Strand | Lenght | Start cod. | Stop cod. | Intergenic nucl[a] |
|---|---|---|---|---|---|---|
| **1-999** | ***nad2*** | **H** | **999** | **ATA** | **T-** | **-2** |
| **998-1062** | ***trnW*** | **H** | **65** | | | **-8** |
| **1055-1129** | ***trnC*** | **J** | **75** | | | **0** |
| **1121-1185** | ***trnY*** | **J** | **65** | | | **1** |
| **1187-2725** | ***coxI*** | **H** | **1539** | **ATG** | **T-** | **-5** |
| **2721-2785** | ***trnL2*** | **H** | **65** | | | **0** |
| **2786-3487** | ***coxII*** | **H** | **702** | **ATA** | **T-** | **-35** |
| **3453-3522** | ***trnK*** | **H** | **70** | | | **-1** |
| **3522-3586** | ***trnD*** | **H** | **65** | | | **0** |
| **3587-3745** | ***atp8*** | **H** | **159** | **ATA** | **TAA** | **-7** |
| **3739-4413** | ***atp6*** | **H** | **675** | **ATG** | **TAA** | **-1** |
| **4413-5237** | ***coxIII*** | **H** | **825** | **ATG** | **T-** | **-38** |
| **5200-5264** | ***trnG*** | **H** | **65** | | | **0** |
| **5265-5630** | ***nad3*** | **H** | **366** | **ATA** | **T-** | **-14** |
| **5617-5683** | ***trnA*** | **H** | **67** | | | **-1** |
| **5683-5757** | ***trnR*** | **H** | **75** | | | **0** |
| **5758-5823** | ***trnN*** | **H** | **66** | | | **1** |
| **5825-5879** | ***trnS1*** | **H** | **55** | | | **1** |
| **5891-5945** | ***trnE*** | **H** | **55** | | | **-2** |
| **5944-6007** | ***trnF*** | **J** | | | | **-1** |
| **6007-7671** | ***nad5*** | **J** | **1665** | **ATA** | **TAA** | **60** |
| **7732-7795** | ***trnH*** | **J** | **64** | | | **-10** |
| **7786-9125** | ***nad4*** | **J** | **1338** | **ATG** | **TAA** | **-7** |
| **9119-9403** | ***nad4L*** | **J** | **285** | **ATA** | **TAA** | **8** |
| **9412-9476** | ***trnT*** | **H** | **65** | | | **0** |
| **9476-9540** | ***trnP*** | **J** | **65** | | | **1** |
| **9542-10021** | ***nad6*** | **H** | **480** | **ATA** | **TAA** | **-1** |
| **10021-11154** | ***cob*** | **H** | **1134** | **ATG** | **T-** | **-2** |
| **11153-11219** | ***trnS2*** | **H** | **67** | | | **-65** |
| **11155-12183** | ***nad1*** | **J** | **1029** | **ATA** | **TAG** | **3** |
| **12187-12253** | ***trnL1*** | **J** | **64** | | | **0** |
| **12254-13532** | ***rrnL*** | **J** | **1279** | | | **0** |
| **13533-13601** | ***trnV*** | **J** | **69** | | | **0** |
| **13602-14152** | ***rrnS*** | **J** | **551** | | | |

Note.- [a] Negatives numbers indicate that adjacent genes overlap.

Tab9: Organization of *Bacillus atticus* mitochondrial genome.

| Position | Gene | Strand | Lenght | Start c. | Stop c. | Intergenic nucl[a] |
|---|---|---|---|---|---|---|
| **1-999** | ***nad2*** | **H** | **999** | **ATA** | **T-** | **-2** |
| **998-1062** | ***trnW*** | **H** | **65** | | | **-7** |
| **1055-1119** | ***trnC*** | **J** | **65** | | | **0** |
| **1120-1183** | ***trnY*** | **J** | **64** | | | **1** |
| **1185-2723** | ***coxI*** | **H** | **1539** | **ATG** | **T-** | **-5** |
| **2719-2782** | ***trnL2*** | **H** | **64** | | | **0** |
| **2783-3484** | ***coxII*** | **H** | **702** | **ATA** | **T-** | **-35** |
| **3450-3519** | ***trnK*** | **H** | **70** | | | **-1** |
| **3519-3584** | ***trnD*** | **H** | **66** | | | **0** |
| **3585-3743** | ***atp8*** | **H** | **159** | **ATA** | **TAA** | **-7** |
| **3737-4411** | ***atp6*** | **H** | **675** | **ATG** | **TAA** | **-1** |
| **4411-5235** | ***coxIII*** | **H** | **825** | **ATG** | **T-** | **-38** |
| **5198-5262** | ***trnG*** | **H** | **65** | | | **0** |
| **5263-5628** | ***nad3*** | **H** | **366** | **ATA** | **T-** | **-14** |
| **5615-5691** | ***trnA*** | **H** | **77** | | | **-1** |
| **5691-5746** | ***trnR*** | **H** | **56** | | | **0** |
| **5747-5812** | ***trnN*** | **H** | **66** | | | **1** |
| **5814-5878** | ***trnS1*** | **H** | **65** | | | **1** |
| **5880-5945** | ***trnE*** | **H** | **66** | | | **-2** |
| **5944-6007** | ***trnF*** | **J** | **64** | | | **-1** |
| **6007-7671** | ***nad5*** | **J** | **1665** | **ATG** | **TAA** | **60** |
| **7732-7795** | ***trnH*** | **J** | **64** | | | **2** |
| **7798-9133** | ***nad4*** | **J** | **1335** | **ATG** | **TAA** | **-7** |
| **9116-9401** | ***nad4L*** | **J** | **285** | **ATA** | **TAA** | **8** |
| **9410-9474** | ***trnT*** | **H** | **65** | | | **0** |
| **9475-9538** | ***trnP*** | **J** | **64** | | | **1** |
| **9540-10019** | ***nad6*** | **H** | **480** | **ATA** | **TAA** | **-1** |
| **10019-11146** | ***cob*** | **H** | **1128** | **ATG** | **T-** | **-2** |
| **11145-11211** | ***trnS2*** | **H** | **67** | | | **-67** |
| **11147-12175** | ***nad1*** | **J** | **1029** | **ATA** | **TAG** | **3** |
| **12179-12255** | ***trnL1*** | **J** | **77** | | | **0** |
| **12256-13536** | ***rrnL*** | **J** | **1281** | | | **0** |
| **13537-13595** | ***trnV*** | **J** | **59** | | | **0** |
| **13596-14138** | ***rrnS*** | **J** | **543** | | | |

Note.- [a] Negatives numbers indicate that adjacent genes overlap.

Table10: Genbank accession numbers for taxa used in this study.

| Order | Species | Genebank acces no. | Reference |
|---|---|---|---|
| **Archaeognatha** | ***Nesomachilis australica*** | NC_006895 | Cameron et al., 2004 |
| **Thysanura** | ***Tricholepidion gertischi*** | NC_005437 | Nardi et al., 2003 |
| **Ephemeroptera** | ***Parafronurus youi*** | NC_011359 | Zhang et al., 2008 |
| **Odonata** | ***Orthetrum triangolare*** | AB126005 | Yamauchi et al., 2004 |
| **Blattaria** | ***Periplaneta fuliginosa*** | NC_006076 | Yamauchi et al., 2004 |
| **Isoptera** | ***Reticulitermes hageni*** | NC_009501 | Cameron and Whiting, 2007 |
| **Mantodea** | ***Tamolanica tamolana*** | NC_007702 | Cameron et al., 2006 |
| **Orthoptera** | ***Locusta migratoria*** | NC_001712 | Flook et al., 1995b |
| | ***Grillotalpa orientalis*** | NC_006678 | Kim et al., 2005 |
| **Mantophasmatodea** | ***Sclerophasma paresisensis*** | NC_007701 | Cameron et al., 2006 |
| **Phasmatodea** | ***Timema californicum*** | DQ241799 | Cameron et al., 2006 |
| | ***Bacillus rossius*** | // | this study |
| | ***Bacillus atticus*** | // | this study |
| **Grylloblattodea** | ***Grylloblatta sculleni*** | DQ241796 | Cameron and Whiting, 2007 |

## 7. FIGURES

Figure 1.

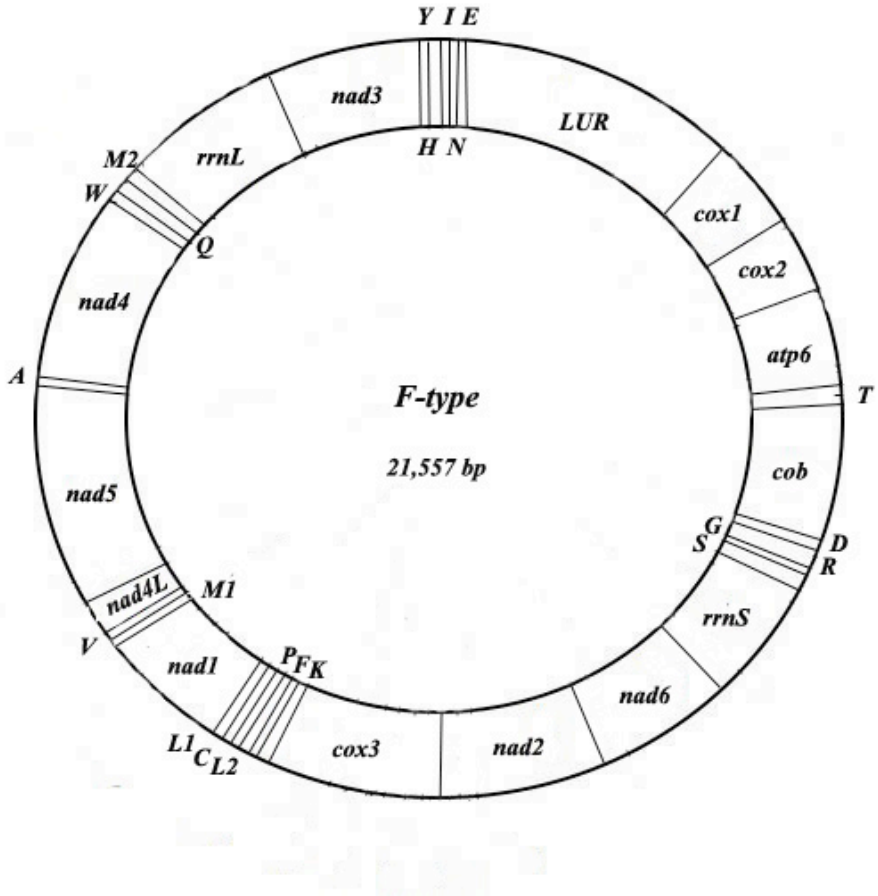


Figure 1: Gene map of the female *Musculista senhousia* mitochondrial genome.

Figure 2.

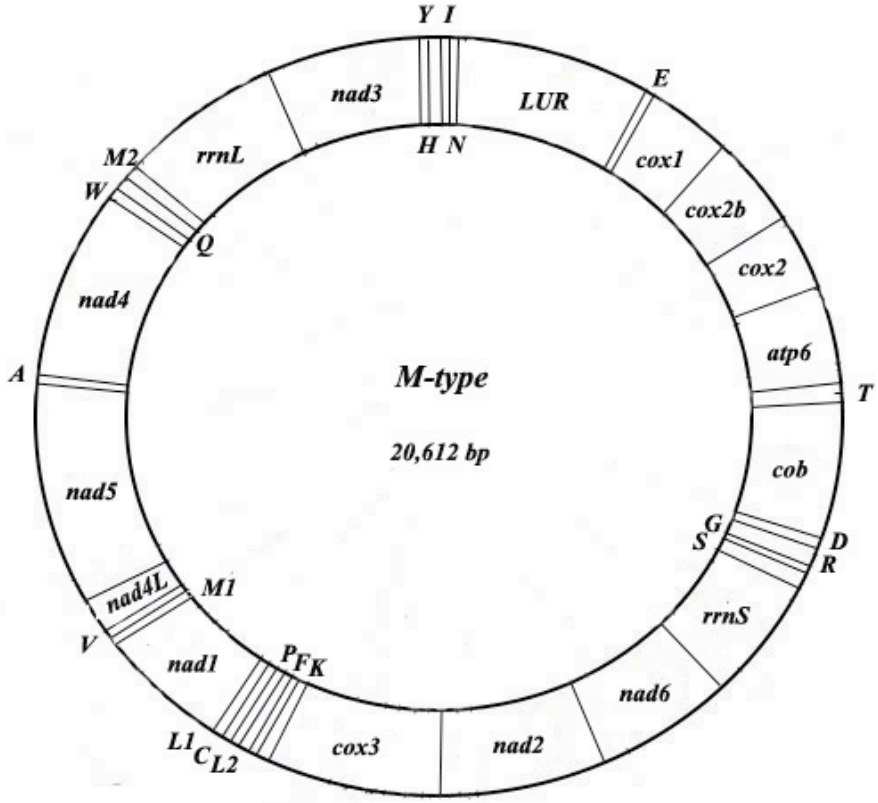


Figure 2: Gene map of the male *Musculista senhousia* mitochondrial genome.

Figure 3.

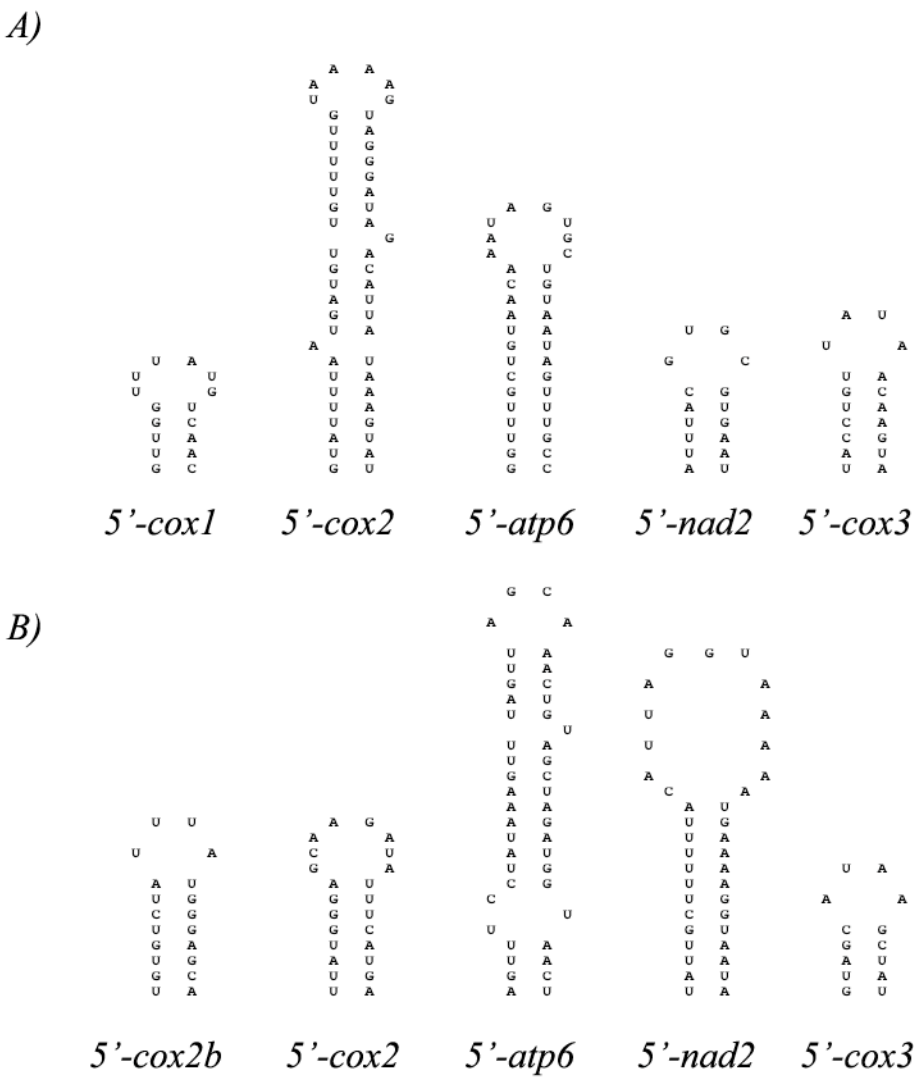


Figure 3: Putative secondary structure preceding the 5'-end of protein-coding genes.

A): Female *Musculista senhousia* mitochondrial genome.

B): Male *Musculista senhousia* mitochondrial genome.

Figure 4.

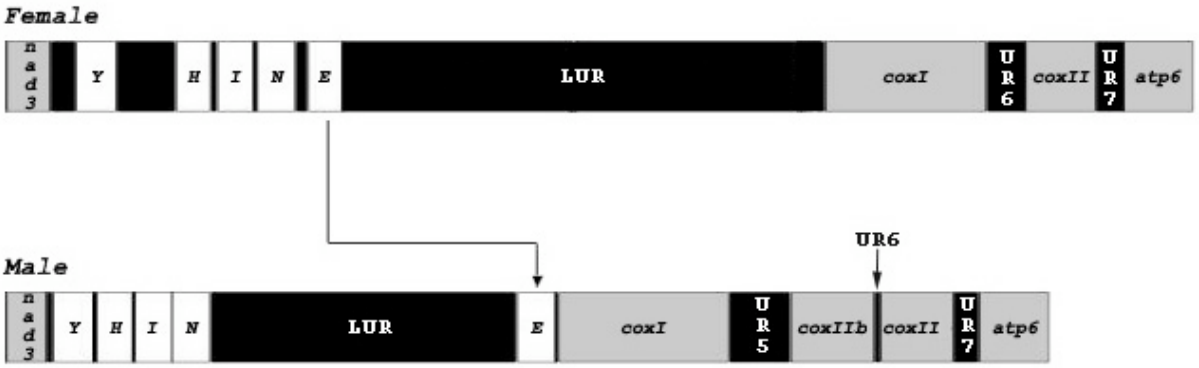


Figure 4: particular of the section relative to the *coxII* gene duplication in male *Musculista senhousia* mitochondrial genome.

Figure 5.

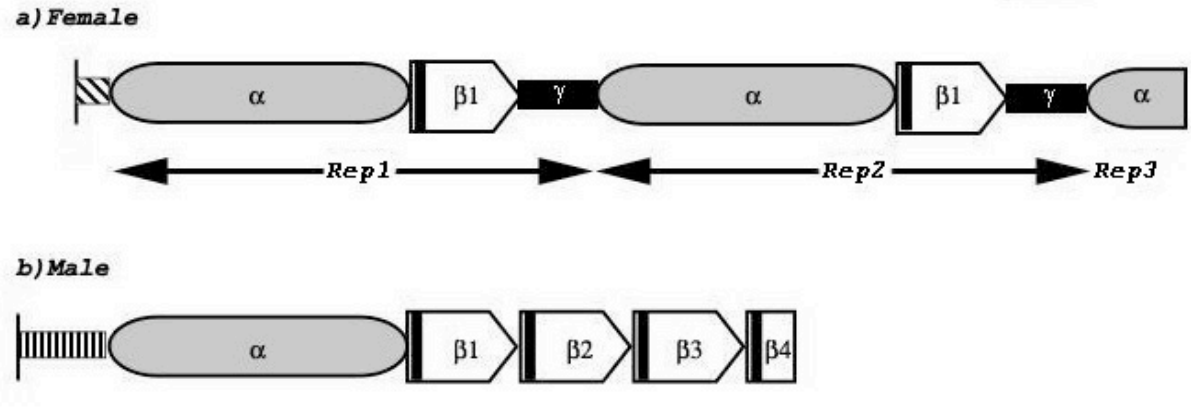


Figure 5: Schematic structure of both a) male and b) female LUR (Large Unassigned Region) of *Musculista senhousia.* Within β repeats, black vertical solid line indicates the position of palindromic sequence.

Figure 6.

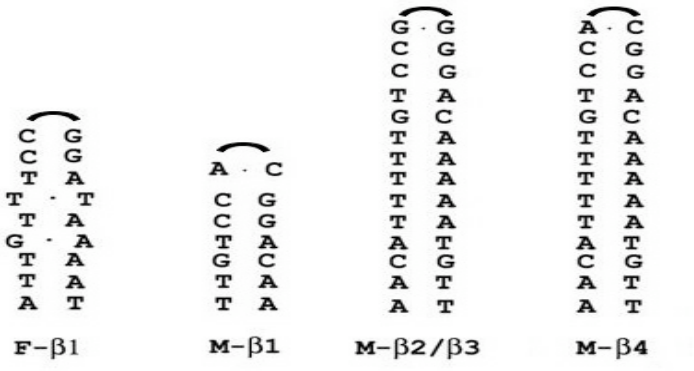


Figure 6: Sequences and structures of palindromic motifs located within β repeats.

Figure7.

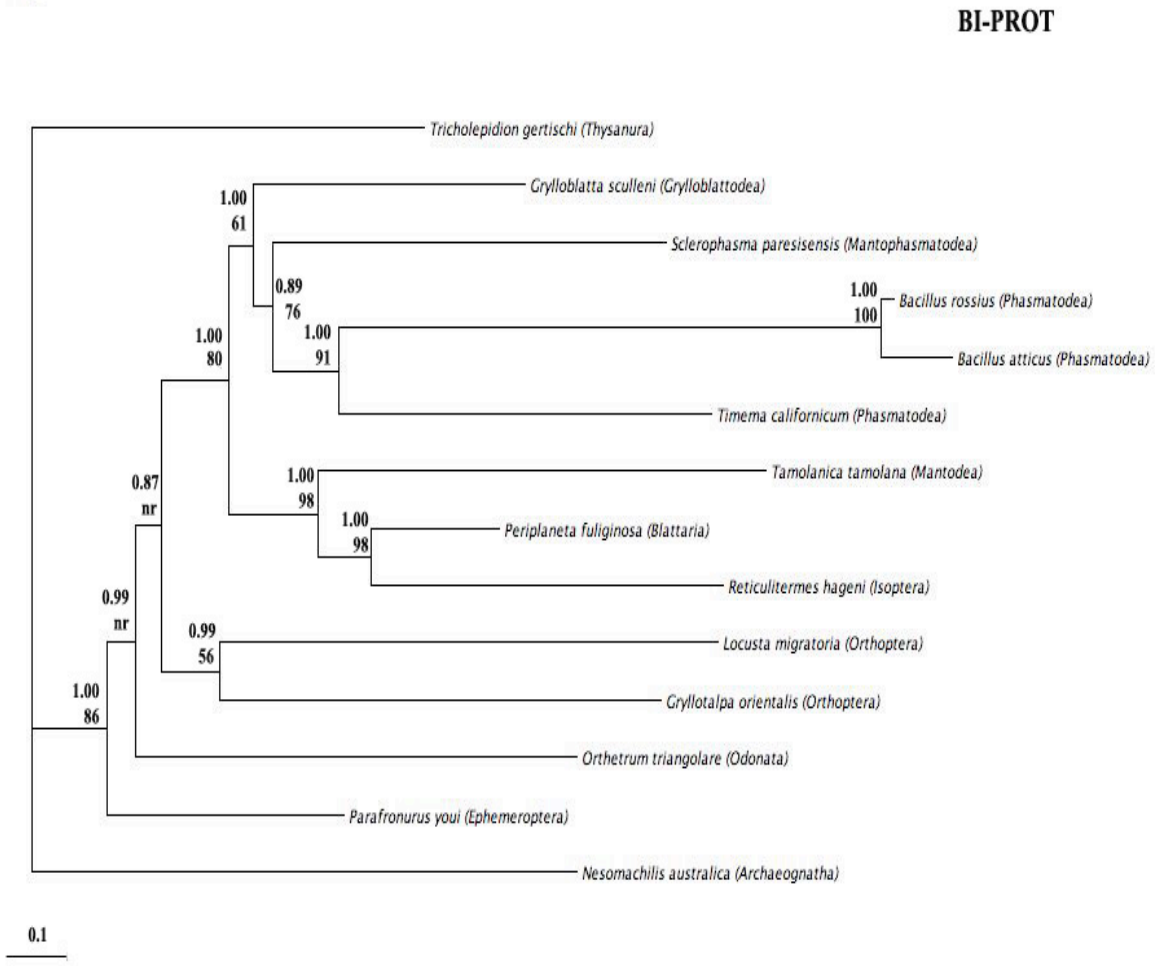


Figure 7: Phylogenetic tree of the relationships among *Bacillus* and other 12 taxa based on the amino acid dataset of 13 mitochondrial protein-coding genes. Branch lengths and topologies from Bayesian analyses. Numbers above branches specify posterior probabilities from Bayesian inference (BI) and bootstrap percentages from maximum parsimony (MP) analyses.

Figure 8.

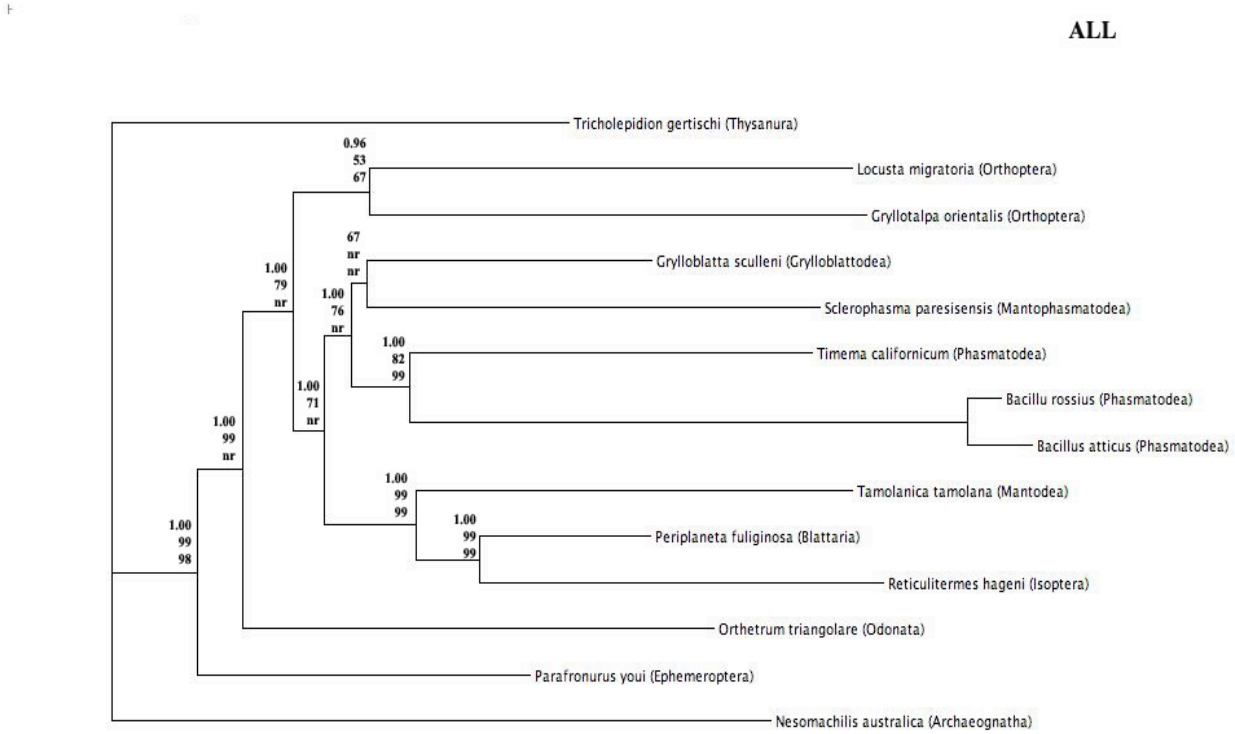


Figure 8: Phylogenetic tree of the relationships among *Bacillus* and 12 other taxa based on the total nucleotide dataset (ALL). Branch lengths and topology are from the Bayesian analysis. Numbers above branches specify posterior probabilities from Bayesian inference (BI), bootstrap percentages from maximum likelihood (ML, 500 replicates) and maximum parsimony (MP, 1500 replicates) analyses.